DOCUMENT RESUME

ED 451 227 TM 032 455

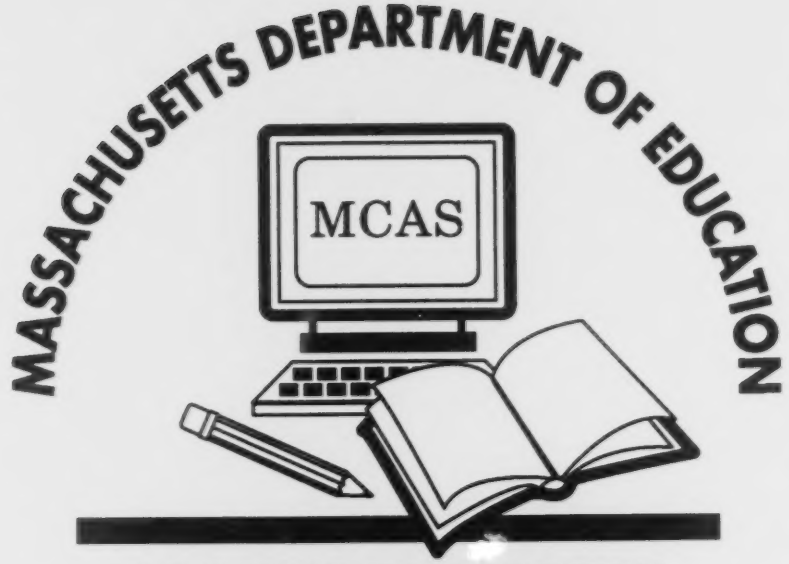
TITLE	Massachusetts Comprehensive Assessment System 1998 Technical Report.
INSTITUTION	Massachusetts State Dept. of Education, Boston.
PUB DATE	1999-10-00
NOTE	175p.
PUB TYPE	Guides - Non-Classroom (055) -- Numerical/Quantitative Data (110)
EDRS PRICE	MF01/PC07 Plus Postage.
DESCRIPTORS	Elementary Education; *Elementary School Students; English; Language Arts; Mathematics; *Psychometrics; Reliability; Sciences; *Scoring; State Programs; Statistical Analysis; *Test Construction; Test Interpretation; *Testing Programs; *Validity
IDENTIFIERS	Massachusetts; *Massachusetts Comprehensive Assessment System

ABSTRACT

This manual documents the technical aspects of the Massachusetts Comprehensive Assessment System (MCAS). In May 1998, Massachusetts public school students in grades 4, 8, and 10 participated in the first annual administration of the MCAS tests in English Language Arts, Mathematics, and Science & Technology. This report provides information about the technical quality of these assessments. It includes a description of the processes used to develop, administer, and score the tests and to analyze test results. The report may be of interest to the educated public, but it is intended for experts in psychometrics and educational research and assumes working knowledge of measurement concepts such as reliability and validity and statistical concepts such as correlation and central tendency. The report contains these sections: (1) "Assessment Development"; (2) "Test Administration"; (3) "Development and Reporting of Scores"; (4) "Technical Characteristics"; (5) "Reference"; and (6) "Appendices." The appendixes list MCAS committee members, and discuss distributions of item statistics and estimating the accuracy of MCAS performance level decisions. (Contains 4 figures, 45 tables, and 24 references.) (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 451 227



MASSACHUSETTS COMPREHENSIVE ASSESSMENT SYSTEM

1998 TECHNICAL REPORT

TM032455

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

H. Reynolds, Jr.

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

OCTOBER 1999

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

CONTENTS

Chapter 1: Background and Overview 1

SECTION I: ASSESSMENT DEVELOPMENT 8
Chapter 2: Overview of Test Design 9
Chapter 3: Design of the English Language Arts Assessment 14
Chapter 4: Design of the Mathematics Assessment 21
Chapter 5: Design of the Science and Technology Assessment 26
Chapter 6: Test Development Process 30

SECTION II: TEST ADMINISTRATION 40
Chapter 7: Test Administration 41

SECTION III: DEVELOPMENT AND REPORTING OF SCORES 48
Chapter 8: Scoring 49
Chapter 9: Standard Setting 57
Chapter 10: Scaling 66
Chapter 11: Score Reporting 68
Chapter 12: State Results 72

SECTION IV: TECHNICAL CHARACTERISTICS 77
Chapter 13: Item Analyses 78
Chapter 14: Reliability 83
Chapter 15: Validity 89

SECTION V: REFERENCES 92

SECTION VI: APPENDICES 95
Appendix A: MCAS Committee Memberships
Appendix B: Distributions of Item Statistics
Appendix C: Estimating the Accuracy and Consistency of MCAS Performance Level Decisions

CHAPTER 1
BACKGROUND AND OVERVIEW

PURPOSE OF THIS MANUAL

The purpose of this technical manual is to document the technical aspect of the Massachusetts Comprehensive Assessment System (MCAS). In May 1998, Massachusetts public school students in grades 4, 8, and 10 participated in the first annual administration of the MCAS tests in English Language Arts, Mathematics, and Science & Technology. This report provides information about the technical quality of those assessments. This includes a description of the processes used to develop, administer, and score the tests and to analyze the test results. This report will serve as a guide for replicating and/or improving the procedures in subsequent years.

Although some parts of this technical report may be used by educated laypersons, the intended audience is experts in psychometrics and educational research. The report assumes working knowledge of measurement concepts such as reliability and validity, and statistical concepts such as correlation and central tendency. For some chapters, the reader is presumed to have basic familiarity with advanced topics in measurement and statistics.

THE EDUCATION REFORM LAW OF MASSACHUSETTS OF 1993

The Massachusetts Comprehensive Assessment System (MCAS) was developed in response to the Education Reform Law of Massachusetts of 1993. Three sections of the reform act that are particularly relevant to the assessment program are restated below.

> *The board shall direct the commissioner to institute a process to develop academic standards for the core subjects of mathematics, science and technology, history and social science, English, foreign languages and the arts. The standards shall cover grades kindergarten through twelve and shall clearly set forth the skills, competencies and knowledge expected to be possessed by all students at the conclusion of individual grades or clusters of grades. The standards shall be formulated so as to set high expectations of student performance and to provide clear and specific examples that embody and reflect these high expectations, and shall be constructed with due regard to the work and recommendations of national organizations, to the best of similar efforts in other states, and to the level of skills, competencies and knowledge possessed by typical students in the most educationally advanced nations. The skills, competencies and knowledge set*

forth in the standards shall be expressed in terms which lend themselves to objective measurement, define the performance outcomes expected of both students directly entering the work force and of students pursuing higher education, and facilitate comparisons with students of other states and other nations.

The "competency determinations" shall be based on the academic standards and curriculum frameworks for tenth graders in the areas of mathematics, science and technology, history and social science, foreign languages, and English, and shall represent a determination that a particular student has demonstrated mastery of a common core of skills, competencies and knowledge in these areas, as measured by the assessment instruments described in section one I. Satisfaction of the requirements of the competency determination shall be a condition for high school graduation. If the particular student's assessment results for the tenth grade do not demonstrate the required level of competency, the student shall have the right to participate in the assessments program the following year or years.

... comprehensive diagnostic assessment of individual students shall be conducted at least in the fourth, eighth and tenth grades. Said diagnostic assessments shall identify academic achievement levels of all students in order to inform teachers, parents, administrators and the students themselves, as to individual academic performance. The board shall develop procedures for updating, improving or refining the assessment system. The assessment instruments shall be designed to avoid gender, cultural, ethnic or racial stereotypes and shall recognize sensitivity to different learning styles and impediments to learning. The system shall take into account on a nondiscriminatory basis the cultural and language diversity of students in the commonwealth and the particular circumstances of students with special needs. Said system shall comply with federal requirements for accommodating children with special needs. All potential English proficient students from language groups in which programs of transitional bilingual education are offered under chapter seventy-one A shall also be allowed opportunities for assessment of their performance in the language which best allows them to demonstrate educational achievement and mastery. For the purposes of this section, a "potential English proficient student" shall be defined as a student who is not able to perform ordinary class work in English; provided, however, that no student shall be allowed to be tested in a language other than English for longer than three consecutive years.

CURRICULUM FRAMEWORKS

As required by the Educational Reform Act of 1993, the Massachusetts Department of Education developed and disseminated *Curriculum Frameworks*. These frameworks are intended to provide guidance for the reform of public education in Massachusetts by raising the standards and expectations of schools and students. The following three frameworks guided the development of MCAS test specifications (Massachusetts Department of Education, 1997a, 1997b, 1997c):

- *English Language Arts Curriculum Framework,*
- *Mathematics Curriculum Framework: Achieving Mathematical Power;* and
- *Science and Technology Curriculum Framework: Owning the Questions through Science and Technology.*

English Language Arts

The English language arts standards are divided into four strands: language, literature, composition, and media. The framework also provides two suggested lists of authors, illustrators, and works.

Mathematics

The mathematics standards are divided into four content-based strands: number sense; patterns, relations, and functions; geometry and measurement; and statistics and probability. The framework also discusses four aspects of applying mathematical knowledge: problem solving, communication, reasoning, and connections.

Science & Technology

The science and technology standards are divided into four strands: inquiry; domains of science; technology; and science, technology, and human affairs. Domains of science is divided into three substrands: physical sciences, life sciences, and earth and space sciences. Technology is divided into two substrands: the design process and understanding and using technology.

PURPOSES OF THE MCAS

The statewide assessment program serves two main purposes. First, it is an accountability tool for measuring the performance of individual students, schools and districts against established state standards. Second, it is intended to improve classroom instruction by a) providing useful feedback about the quality of instruction and b) modeling effective assessment approaches that can be used in the classroom.

The Education Reform Law requires that students demonstrate competency on the tenth grade MCAS tests. In addition to fulfilling local graduation requirements, students must pass the state's grade 10 tests as a condition for receiving a high school diploma. The Massachusetts Board of

Education has determined that this requirement will be applied for the first time to graduates of the Class of 2003. Students will be given multiple opportunities, if necessary, to pass the tests. In the future, the Board of Education will determine the standard for passing the MCAS grade 10 tests.

The Education Reform Law also requires the Department of Education to evaluate whether schools and districts are improving students' performance based on the learning standards contained in the *Curriculum Frameworks*. Once in place, this evaluation of school and district performance will be based in part on results from the MCAS tests.

Local educators should use results of the MCAS tests, together with results of local tests and assessments, to identify strengths and weaknesses in curriculum and instruction, and to determine the needs of individual students in order to serve them more effectively. As part of the MCAS results, local educators should make use of released MCAS test items, *The Massachusetts Comprehensive Assessment System Release of May 1998 Test Items (1998a)*, and the *Test Item Analysis Report* (which contains student results for each of the questions provided in that year's release document). These resources, along with other resources provided by the Department of Education, can assist educators in developing and implementing instructional strategies designed to support the goal that all students attain the state's academic learning standards.

ORGANIZATION OF THIS MANUAL

The organization of this report is based on the conceptual flow of an assessment's life span; it begins with the initial test specification and addresses all the intermediate steps that lead to final score reporting. Section I covers the development of the MCAS tests. It consists of five chapters, covering general design issues, the specific designs of the English Language Arts, Mathematics, and Science & Technology assessments, and the test development process. Section II consists of one chapter describing the administration of the tests. Section III contains five chapters covering scoring, standard setting, scaling, score reporting, and state results. Section IV presents three chapters addressing the technical characteristics of the tests. Topics covered include item analysis, reliability, and validity.

Because of the educational and political importance of high-stakes testing programs such as the MCAS, this technical report uses professional guidelines for evaluating and documenting the testing program, specifically the *Standards for Educational and Psychological Testing* (AERA, APA, and NCME, 1985) and the *Code of Fair Testing Practices in Education* (1988). The *Standards for Educational and Psychological Testing* covers technical standards for test development and evaluation, professional standards for test use, standards for particular applications (i.e., testing students of limited English proficiency and students with disabilities), and standards for administrative procedures (i.e., test administration, scoring and reporting, and protecting the rights of test takers). Table 1-1 shows the categories of standards from the *Standards for Educational and Psychological Testing* and shows where each category of standards is addressed in this technical manual report or elsewhere.

Table 1-1
Information Addressing Standards in the *Standards for Educational and Psychological Testing*

Standards		Location of Information
Technical Standards for Test Construction and Evaluation	Validity	Chapter 15
	Reliability and Errors of Measurement	Chapter 14
	Test Development and Revision	Chapters 2–6
	Scaling, Norming, Score Comparability, and Equating	Chapter 10 (Scaling, other topics not applicable)
	Test Publication: Technical Manuals and User's Guides	Chapters 1–15
Professional Standards for Test Use	General Principals of Test Use	Throughout technical manual
	Clinical Testing	Not applicable
	Educational and Psychological Testing in the Schools	Throughout technical manual
	Test Use in Counseling	Not applicable
	Employment Testing	Not applicable
	Professional and Occupational Licensure and Certification	Not applicable
	Program Evaluation	Not applicable for 1998 test
Standards for Particular Applications	Testing Linguistic Minorities	Chapter 7
	Testing People Who Have Handicapping Conditions	Chapter 7
Standards for Administrative Procedures	Test Administration, Scoring, and Reporting	Chapters 7, 8, 11
	Protecting the Rights of Test Takers	Not covered in technical manual*

*Addressed in administration manuals prepared for principals and test administrators and also in Requirements for Participation.

The *Code of Fair Testing Practices in Education* covers developing appropriate tests, interpreting scores, striving for fairness, and informing test takers. Table 1-2 shows where each point covered by the *Code of Fair Testing Practices in Education* is addressed.

Table 1-2
Information Regarding Responsibilities for Test Developers in
Code of Fair Testing Practices in Education

	Responsibility	Location of Information
Developing Appropriate Tests	Define what each test measures and what the test should be used for. Describe the populations for which the test is appropriate.	Chapters 1–5, 7; *MCAS Guides;* Special Education Advisory; Requirements for Participation
	Accurately represent the characteristics, usefulness, and limitations of each test for its intended purposes.	Chapter 2; *MCAS Guides; Guide to Interpreting the 1998 MCAS School and District Reports*
	Explain relevant measurement concepts as necessary for clarity at the level of detail that is appropriate for the intended audiences.	Chapters 9, 10, 13–15
	Describe the process of test development. Explain how the content and skills to be tested were selected.	Chapter 3–6
	Provide evidence that the test meets its intended purpose(s).	Chapters 2–5, 15
	Provide representative samples or complete copies of test questions, directions, answer sheets, manuals, and score reports to qualified users.	Chapter 11; *Release of May 1998 Test Items,* Item tryouts, administration manuals
	Indicate the nature of the evidence obtained concerning the appropriateness of each test for groups of different racial, ethnic, or linguistic backgrounds who are likely to be tested.	Chapter 13, *Bias Review*
	Identify and publish any specialized skills needed to administer each test and to interpret scores correctly.	Not Applicable
Interpreting Scores	Provide timely and easily understood score reports that describe test performance clearly and accurately. Also explain the meaning and limitations of reported scores.	Chapter 11
	Describe the population(s) represented by any norms or comparison group(s), the dates the data were gathered, and the process used to select the samples of test takers.	Chapter 7
	Warn users to avoid specific, reasonably anticipated misuses of test scores.	*Guide to Interpreting the 1998 MCAS School and District Reports, Understanding Your MCAS 1998 Student Report for Parents/Guardians*
	Provide information that will help users follow reasonable procedures for setting passing scores when it is appropriate to use such scores with the test.	Chapter 9

Table 1-2 Information Regarding Responsibilities for Test Developers in *Code of Fair Testing Practices in Education*		
Responsibility		Location of Information
	Provide information that will help users gather evidence to show that the test is meeting its intended purpose(s).	Chapters 2–5, 15
Striving for Fairness	Review and revise test questions and related materials to avoid potentially insensitive content or language.	Chapter 6
	Investigate the performance of test takers of different races, genders, and ethnic backgrounds when samples of sufficient size are available. Enact procedures that help to ensure that differences in performance are related primarily to the skills under assessment rather than to irrelevant factors.	Chapters 6, 13, *Bias Review*
	When feasible, make appropriately modified forms of tests or administration procedures available for test takers with handicapping conditions. Warn test users of potential problems in using standard norms with modified tests or administration procedures that result in noncomparable scores.	Chapter 7
Informing Test Takers	When a test is optional, provide test takers or their parents/guardians with information to help them judge whether the test should be taken, or if an available alternative to the test should be used.	Not Applicable
	Provide test takers the information they need to be familiar with the coverage of the test, the types of question formats, the directions, and appropriate test-taking strategies. Strive to make such information equally available to all test takers.	*MCAS Guides,* Item Tryouts, Practice Tests, Administration Manuals, DOE Web Site.
	Provide test takers or their parents/guardians with information about rights test takers may have to obtain copies of tests and completed answer sheets, retake tests, have tests rescored, or cancel scores.	*Test Item Analysis Report* and Appeals Policy planned for 1999.
	Tell test takers or their parents/guardians how long scores will be kept on file and indicate to whom and under what circumstances test scores will or will not be released.	Administration manuals and *Understanding Your MCAS 1998 Student Report for Parents/Guardians*
	Describe the procedures that test takers or their parents/guardians may use to register complaints and have problems resolved.	Public outreach campaign and MCAS Support Services center

Despite the many pages of tables, figures, and text in this manual, it is beyond the scope of this report to provide all available details about the MCAS. However, details that are pertinent to understanding the technical quality of the MCAS are included in the appendices or referenced in this manual.

SECTION I
ASSESSMENT DEVELOPMENT

CHAPTER 2
OVERVIEW OF TEST DESIGN

According to the *Standards of Educational and Psychological Testing* (1985, p. 9), the construct that a test is intended to measure should be embedded in a conceptual framework. This chapter discusses the conceptual framework that was used to design the MCAS assessments. The *Standards* (1985) also states (p. 25) that specifications used in constructing the test should be stated clearly. This chapter describes the specifications used for test construction. The MCAS test design and content covered has been explicated previously in two sets of documents: The *Curriculum Frameworks*, which present the learning standards intended to guide the development of local curriculum, and the *Guides to the Massachusetts Comprehensive Assessment System*, which describe what will be on the test. This chapter will summarize pertinent information from those two sets of materials and provide some additional detail.

GUIDES TO THE MASSACHUSETTS COMPREHENSIVE ASSESSMENT SYSTEM

The Education Reform Law of Massachusetts stipulates that the MCAS be based on *the Curriculum Frameworks* for English language arts, mathematics, and science and technology. The Department of Education convened committees of educators[1] from around the state to work with the Department and its testing contractor to design and develop assessments of the learning standards contained in the *Curriculum Frameworks.*

To design the assessments, the *Curriculum Frameworks* were evaluated to determine for each subject area which dimensions could be adequately assessed in an on-demand paper-and-pencil test. A product of this process was the *Guide to the Massachusetts Comprehensive Assessment System*[2] for each test (here called the *MCAS Guides).* The *MCAS Guides* provided the foundation for the test

---

1 Members of different MCAS committees are listed in Appendix A.

2 Massachusetts Department of Education (1998b), *Guide to the Massachusetts Comprehensive Assessment System: English Language Arts,* Malden.

Massachusetts Department of Education (1998c), *Guide to the Massachusetts Comprehensive Assessment System: Mathematics.*

Massachusetts Department of Education (1998d), *Guide to the Massachusetts Comprehensive Assessment System: Science and Technology.*

specifications that detail what each test will cover and emphasize, including the content strands (subject areas) and question types to be used in MCAS.

ITEM TYPES

Every item type has its strengths and weaknesses. To ensure the strongest possible program, each MCAS test used one or more of four different item types: multiple-choice, short answer, open response, and writing prompt.

Multiple-choice questions are highly efficient in terms of testing time, and thus allow for a breadth of content coverage. Multiple-choice questions, however, are susceptible to guessing and, for tests requiring computation (much of mathematics and for some aspects of science), are susceptible to back solving. That is, instead of using the intended solution strategy, students can insert each choice into the problem and rule out incorrect options, one by one. MCAS multiple-choice items were scored one point if correct and zero points if incorrect.

Short-answer questions require responses ranging from a few words or a number to several sentences. They are relatively immune to random guessing and back solving. For these reasons, MCAS used short-answer questions as part of the mathematics assessment. MCAS short-answer items were scored on a 0–1 scale.

Open-response (extended-response) questions invite students to demonstrate not only their knowledge of facts and comprehension about a subject, but also how they can apply their knowledge. Open-response questions can take many forms, but they all require students to construct a detailed or descriptive answer (usually up to half a page long), and take between ten and fifteen minutes to complete. MCAS open-response questions were all scored on a 0–4 scale.

MCAS writing prompts require students to write one or more pieces, which are then evaluated by human scorers. Features of the MCAS writing prompts are described in Chapter 3 (in the section titled "Composition"), and scoring of the writing prompts is discussed in Chapter 8.

COMMON-MATRIX DESIGN

MCAS test questions are assigned to either the common or matrix-sampled portions of the tests. Common test questions are those that were identical in all twelve forms of the test at each grade level. Approximately eighty percent of the questions on any given test form were common questions. All individual student results (performance levels, scaled scores, subject subarea information) are based exclusively on common questions; thus, the performance of every student at a grade level is based on identical questions. In addition, performance level results and average scaled scores for schools and districts are based exclusively on common questions.

The remaining twenty percent of the MCAS test questions in each test form were matrix-sampled questions, which differed across the twelve test forms at each grade level tested. Matrix-sampled questions serve two primary purposes. First, starting in the second year of the testing program, they will serve as the basis for equating tests from year to year. This allows for comparisons of performance at the school and district levels over time. Second, matrix-sampled questions, when combined with common questions, allow reporting in greater depth and detail for a broader range of the curriculum than is possible with common questions only. Results from the matrix-sampled questions and common questions are aggregated at the school and district levels to produce subject area subscores.

Common questions are publicly released following each year's test administration to inform local decisions about curriculum and instruction.[3] Released common questions are replaced each year with either questions from the previous year's matrix-sampled section or newly developed field-tested questions.

The distribution of common and matrix-sampled questions for each grade level is shown in Table 2-1.

---

[3] Massachusetts Department of Education (1998). *The Massachusetts Comprehensive Assessment System: Release of May 1998 Test Items.*

Table 2-1
May 1998 MCAS
Number of Test Questions in Each Content Area by Question Type and Function

Question Type: MC = Multiple- Choice, SA = Short Answer, OR = Open Response, WP = Writing Prompt

Grade	Question Function	Content Area							
		English Language Arts			Mathematics			Science & Technology	
		MC	OR	WP	MC	SA	OR	MC	OR
4	Common	28	5	1	21	5	6	26	6
	Matrix	8	2	1	5	1	1	6	1
	Total	36	7	2	26	6	7	32	7
8	Common	28	5	1	21	5	6	26	6
	Matrix	8	2	1	5	1	1	6	1
	Total	36	7	2	26	6	7	32	7
10	Common	32	8	1	27	5	7	32	8
	Matrix	8	2	1	7	1	2	8	2
	Total	40	10	2	34	6	9	40	10

TEST SESSION STRUCTURE

Within each subject, test questions were organized in separate 45-minute sessions. The number of questions per session was based on estimated time spent on each type of question. For reading (language and literature), the length of the selection was also factored in. However, Department policy was to provide students with as much time as they could use productively (and without compromising schools' administration constraints). The amount of additional time per session that was generally considered reasonable ranged from five minutes to one-half hour. The number of sessions administered at each grade level in each subject area is shown in Table 2-2.

Table 2-2
Number of 45-Minute Test Sessions
Administered at Each Grade Level by Subject Area

Subject	Grade 4	Grade 8	Grade 10
English Language Arts	7	7	7
Mathematics	3	3	4
Science & Technology	3	3	4
All Subjects	13	13	15

Each test booklet for each grade level included seven separate English language arts sessions (labeled 1, 2, 3A, 3B, 4, 5, and 6). Sessions 1, 4, and 5 included a reading selection, followed by

multiple-choice and open-response questions. All questions in Sessions 1, 4, and 5 were common questions. In Session 3A, students were required to write a draft of a long composition in response to a writing prompt. In Session 3B, students revised the draft of their long composition, producing their final long composition in response to the writing prompt given in Session 3A. A single writing prompt for Sessions 3A and 3B was administered to all students within a grade level. Sessions 2 and 6 were comprised of matrix questions. Session 2 contained both multiple-choice and open-response questions. Session 6 contained the writing prompt for the short composition. In the sessions that contained both multiple-choice and open-response questions, the multiple-choice questions appeared first in the test booklet, followed by the open-response questions.

Each test session in mathematics included multiple-choice, short-answer, and open-response questions, with the exception of Session 3 for grade 4, Session 2 for grade 8, and Sessions 3 and 4 for grade 10, which did not include short-answer questions. Multiple-choice questions appeared first in the test booklet for each session. Next were the open-response and short-answer questions, which were interspersed.

Science & Technology sessions for all grades included multiple-choice and open-response questions only. As in the other tests, multiple-choice questions appeared first in each session, followed by open-response questions.

CHAPTER 3
DESIGN OF THE ENGLISH LANGUAGE ARTS ASSESSMENT

BACKGROUND

The English Language Arts section of the Massachusetts Comprehensive Assessment System (MCAS) is based exclusively on the learning standards described in the Massachusetts *English Language Arts Curriculum Framework* (1997). These learning standards were developed in collaboration with teachers, school and district administrators, reading and writing specialists, college faculty, and parents. The *English Language Arts Curriculum Framework* identifies expectations for student learning for grade groupings Pre-K-4, 5-8, 9-10, and 11-12.

CONTENT STRANDS

Three content strands identified by the *English Language Arts Curriculum Framework* served as the foundation for the MCAS English Language Arts assessment:

- Language
- Literature
- Composition

The MCAS English Language Arts assessment addressed all of the learning standards contained in the *English Language Arts Curriculum Framework* that are feasible to assess in an on-demand test format. Certain learning standards from the Language, Literature, and Composition Strands – for example, learning standard 3, "Students will make oral presentations..." – were not tested on the MCAS English Language Arts test. In addition, all three learning standards of the *Framework*'s Media Strand and were not tested by MCAS.

The *Guide to the Massachusetts Comprehensive Assessment System: English Language Arts* identifies the following standards assessed by the MCAS on-demand tests: language strand 4–7, literature strand 8–17, and composition strand 19–22. Table 3-1 presents the English language arts learning standards from the *English Language Arts Curriculum Framework*.

Table 3-1
English Language Arts Learning Standards

	1	Use agreed-upon rules for informal and formal discussions in small and large groups.
	2	Pose questions, listen to the ideas of others, and contribute their own information or ideas in group discussions and interviews in order to acquire new knowledge.
	3	Make oral presentations that demonstrate appropriate consideration of audience, purpose, and the information to be conveyed.
	4	Acquire and use correctly an advanced reading vocabulary of English words, identifying meanings through an understanding of word relationships.
	5	Identify, describe, and apply knowledge of the structure of the English language and standard English conventions for sentence structure, usage, punctuation, capitalization, and spelling.
	6	
	7	

Table 3-1 English Language Arts Learning Standards		
Literature Strand	8	[illegible]
	9	Identify the basic facts and essential ideas in what they have read, heard, or viewed.
	10	Demonstrate an understanding of the characteristics of different genres.
	11	Identify, analyze, and apply knowledge of theme in literature and provide evidence from the text to support their understanding.
	12	Identify, analyze, and apply knowledge of the structure and elements of fiction and provide evidence from the text to support their understanding.
	13	Identify, analyze, and apply knowledge of the structure, elements, and meaning of nonfiction or informational material and provide evidence from the text to support their meaning.
	14	Identify, analyze, and apply knowledge of the structure, elements, and theme of poetry and provide evidence from the text to support their understanding.
	15	Identify and analyze how an author's choice of words appeals to the senses, creates imagery, suggests mood, and sets tone.
	16	Compare and contrast similar myths and narratives from different cultures and geographic regions.
	17	Interpret the meaning of literary works, nonfiction, films, and media by using different critical lenses and analytic techniques.
	18	Plan and present effective dramatic readings, recitations, and performances that demonstrate appropriate consideration of audience and purpose.
Composition Strand	19	Write compositions with a clear focus, logically related ideas to develop it, and adequate supporting detail.
	20	Select and use appropriate genres, modes of reasoning, and speaking styles when writing for different audiences and rhetorical purposes.
	21	Improve organization, content, paragraph development, level of detail, style, tone, and word choice in revising their compositions.
	22	Use their knowledge of standard English conventions for sentence structure, usage, punctuation, capitalization, and spelling to edit their writing.
	23	Use self-generated questions, note-taking, summarizing, précis writing, and outlining to enhance learning when reading or writing.
	24	Use open-ended research questions, different sources of information, and appropriate research methods to gather information for their research projects.
	25	Develop and use rhetorical, logical, and stylistic criteria for assessing final versions of their compositions or research projects before presenting them to varied audiences.
Media Strand	26	Obtain information by using a variety of media and evaluate the quality of the information obtained.
	27	Explain how techniques used in electronic media modify traditional forms of discourse for different aesthetic and rhetorical purposes.
	28	Design and create coherent media productions with a clear focus, adequate detail, and consideration of audience and purpose.

ASSESSMENT COMPONENTS

There were two components of the MCAS English Language Arts tests:

- Language and Literature
- Composition

Each component used one or more of the following assessment modes:

- multiple-choice
- open-response; and

- writing prompts.

Multiple-choice questions on the MCAS English Language Arts test required students to select the correct answer from a list of four options. Open-response questions (posed only in the Language and Literature Component) required students to create a response. Writing prompts are assignments that direct the student in the creation of a piece of writing.

The number and type of questions (per student) included in each component of the MCAS English Language Arts test are shown in Table 3-2.

Table 3-2
English Language Arts
Distribution of Questions (Number per Student) by Component and Grade Level

Mode of Assessment	Language and Literature Component			Composition Component, Short and Long Sessions		
	Grade 4	Grade 8	Grade 10	Grade 4	Grade 8	Grade 10
Multiple-choice questions	36	36	40	0	0	0
Open-response questions*	8	8	10	0	0	0
Writing prompts	0	0	0	2	2	2

* Open-response questions assess learning standards from the literature strand only.

LANGUAGE AND LITERATURE COMPONENT

The Language and Literature Component of the MCAS English Language Arts test consisted of reading passages followed by related questions that assess learning standards from the Language and Literature Strands of the *English Language Arts Curriculum Framework*. Developmentally appropriate reading passages from a range of literary and informational texts appeared in the Language and Literature Component of MCAS.

READING SELECTIONS

Table 3-3 shows MCAS selections classified by the categories: literary and non-narrative nonfiction.

Table 3-3
Genre of MCAS Selections

Literary	Non-Narrative, Nonfiction *
▪ fiction - poetry - drama ▪ nonfiction - essays - biographies - autobiographies	▪ instructions ▪ informational reports and articles ▪ letters ▪ interviews ▪ reviews ▪ essays ▪ speeches ▪ editorials ▪ critiques

* Emphasis on exposition in earlier grades, moving toward persuasive structures at higher grades.

Table 3-4 Percent of Selections by Genre and Source						
Grade	Literary			Non-Narrative NonFiction		
	Appendix A	Appendix B	Other	Appendix A	Appendix B	Other
4	25	13	12	0	0	50
8	30	15	15	0	0	40
10	30	15	15	5	15	20

COMPOSITION COMPONENT

The Composition Component of the MCAS English Language Arts test included two separate sessions:

- Short Session: one administration of approximately 45 minutes
- Long Session: two consecutive administrations totaling approximately 90 minutes

In each session, students were required to complete a writing assignment in response to a writing prompt. In some cases, the writing prompt was related to a reading passage.

Short Session

The Short Session assessed students' skills at writing for various purposes. The types of writing that were assessed varied by grade level and may have included, as is developmentally appropriate, the following:

- Fiction
- Summaries
- Letters
- Instructions
- Essays
- Comparisons/contrasts
- Descriptions
- Analyses

In the Short Session, students were required to complete the writing assignment in a single test administration; therefore, students' writing samples were treated as "first drafts" in the scoring process. Students were encouraged to organize their thoughts, generate ideas, and make notes in a designated area of the test booklet.

Long Session

The Long Session assessed students' skills at writing in a specific mode. The mode of writing to be assessed at each tested grade level was as follows:

Grade 8: Persuasive writing
Grade 10: Literary analysis

The Long Session was structured to include some of the key elements of the writing process: drafting, revising, and finalizing. Consequently, this session was administered in two consecutive administration periods on the same school day, separated by a short break. In the first administration period, students

BEST COPY AVAILABLE

prepared a first draft of their writing. Students were provided with space in the test booklet to generate and organize ideas and draft their writing. Following the break, students returned to revise and finalize their compositions in the second administration period.

At grade 4, students were asked to produce a piece of narrative writing that chronicled and/or described a particular event or experience. At grade 8, students were asked to take a stand on an issue and write a persuasive essay that would convince the reader to take the same stand. At grade 10, students were required to apply their knowledge of literary elements, themes, and structures by writing an essay that analyzed an excerpt from a literary text.

Table 3-5 list the exact number of items that appeared on the 1998 MCAS English Language Arts tests.

Table 3-5
Distribution of Items, 1998 MCAS English Language Arts Assessment

(MC = Multiple-Choice; OR = Open Response; WP = Writing Prompt)

Grade	Reporting Category	Common			Matrix (Total Across 12 Forms)		
		MC	OR	WP	MC	OR	WP
4	Language*	6	0	0	17	0	0
	Literature	22	5	0	79	24	0
	Composition	0	0	1	0	0	12
	Total	28	5	1	96	24	12
	Language	6	0	0	11	0	0
	Literature	22	5	0	85	24	0
	Composition	0	0	1	0	0	12
	Total	28	5	1	96	24	12
10	Language	6	0	0	16	1	0
	Literature	26	8	0	80	23	0
	Composition	0	0	1	0	0	12
	Total	32	8	1	96	24	12

* In 1998, the grade 4 test included four "stand-alone" language items. These items appeared on the same pages as items associated with reading selections, but were not otherwise linked to the selections.

CHAPTER 4
DESIGN OF THE MATHEMATICS ASSESSMENT

LEARNING STANDARDS

The Mathematics MCAS tests were based exclusively on the learning standards described in the Massachusetts *Mathematics Curriculum Framework* (1996). The *Mathematics Curriculum Framework* identifies expectations for student learning, organized by content strands and substrands for grade groupings K-4, 5-8, 9-10, and 11-12. Table 4-1 presents the mathematics content learning standards for pre-kindergarten through grade 4, grades 5 through 8, and grades 9 and 10.

Table 4-1
Mathematics Learning Standards

	PreK–4	Grades 5–8	Grades 9 and 10
Number Sense	1. Number Sense and Numeration 2. Concepts of Whole Number Operations 3. Fractions and Decimals 4. Estimation 5. Whole Number Computation	1. Number and Number Relationships 2. Number Systems and Number Theory 3. Computation and Estimation 4. Ratio, Proportion, Percent	1. Discrete Mathematics 2. Mathematical Structure 3. Estimation
Patterns, Relations, and Functions	1. Patterns and Relationships 2. Algebra/Mathematical Structures	1. Patterns and Functions 2. Algebra	1. Algebra 2. Functions 3. Trigonometry
Geometry and Measurement	1. Geometry and Spatial Sense 2. Measurement	1. Geometry 2. Measurement 3. Geometric Measurement	1. Geometry and Spatial Sense 2. Measurement 3. Geometry from an Algebraic Perspective
Statistics and Probability	1. Statistics and Probability	1. Statistics 2. Probability	1. Statistics 2. Probability

CONTENT COVERAGE

Table 4-2 presents the approximate percentage of 1998 MCAS mathematics items by content strand.

Content Strand	Grade 4	Grade 8	Grade 10
Number Sense	35%	25%	20%
Patterns, Relations, and Functions	20%	30%	30%
Geometry and Measurement	25%	25%	30%
Statistics and Probability	20%	20%	20%

MATHEMATICAL THINKING SKILLS

In addition to content knowledge, students were expected to demonstrate problem-solving and mathematical communication and reasoning skills, as well as skill at making connections between math content and its real-world application.[1] For the purposes of the MCAS tests, these skills are grouped into three major areas: conceptual understanding, procedural knowledge, and problem solving.

Conceptual Understanding

Items in this area assessed student skills in labeling, verbalizing, and defining concepts; recognizing and generating examples and counter-examples; using models, diagrams, charts, and symbols to represent concepts; translating from one mode of representation to another; and comparing, contrasting, and integrating concepts.

Procedural Knowledge

Items in this area assessed student skills related to executing procedures and verifying results; explaining reasons for steps in procedures; recognizing correct and incorrect procedures; developing new procedures, or extending or modifying familiar ones; and recognizing situations in which a procedure is appropriate, necessary, or correctly applied.

Problem Solving

Items in this area assessed student skills in selecting appropriate mathematical concepts and procedures for real-life and mathematical problem situations and appropriately applying these concepts and procedures; selecting and using appropriate problem-solving strategies; and verifying and generalizing solutions. Table 4-3 presents this information for each grade level.

Mathematical Thinking Skill	Grade 4	Grade 8	Grade 10
[illegible]	40%	30%	30%
Procedural Knowledge	40%	25%	25%
Problem Solving	20%	45%	45%

All questions on the Mathematics tested

- knowledge of learning standards from one or more *Mathematics Curriculum Framework* content strands, and
- one of more mathematical thinking skills.

---

1 The core concept of the Massachusetts *Mathematics Curriculum Framework* "is that students develop mathematical power through problem solving, communication, reasoning and [making] connections" (p. 1).

ITEM TYPES

Students were required to answer items that assess the content knowledge and mathematical thinking skills described below as is developmentally appropriate for each grade level. Three types of mathematics questions were used at each grade level tested:

- multiple-choice
- short answer; and
- open response.

Multiple-choice questions on the MCAS Mathematics tests required students to select the correct answer from a list of four options. Short-answer items required a brief response, usually a short statement or numeric solution to a computation or simple problem. Open-response items required students to show their work in solving a problem and require responses in writing or in the form of a chart, table, diagram, or graph, as appropriate.

The approximate distribution of MCAS mathematics test items by type for each grade is shown in Table 4-4.

		Approximate Number of Test Questions (per student test booklet)
4 and 8		26
		6
	Open response	7
10	Multiple-choice	32
	Short answer	6
	Open response	10

Table 4-5 shows the exact number of items appearing in the 1998 MCAS Mathematics Assessment.

Table 4-5
Distribution of Items, 1998 MCAS Mathematics Assessment

(MC = Multiple-Choice; SA = Short Answer; OR = Open Response)

	Reporting Category	Common			Matrix (Total Across 12 Forms)		
		MC	SA	OR	MC	SA	OR
4		9	2	1	23	5	4
	Patterns, Numbers, and Relations	3	2	2	12	3	3
	Geometry and Measurement	5	1	2	13	4	3
	Statistics and Probability	4	0	1	12	0	2
	Total	21	5	6	60	12	12
8		6	2	1	17	7	3
	Patterns, Numbers, and Relations	5	2	2	14	4	3
	Geometry and Measurement	6	1	2	22	1	1
	Statistics and Probability	4	0	1	7	0	5
	Total	21	5	6	60	12	12
10		7	1	2	17	4	5
	Patterns, Numbers, and Relations	6	1	2	28	2	7
	Geometry and Measurement	8	2	3	27	3	7
	Statistics and Probability	7	1	1	12	3	5
	Total	28	5	8	84	12	24

Calculator use

Students at grades 8 and 10 participated in two MCAS Mathematics test sessions in 1998: One session allowed the use of calculators; the other session required students to compute "by hand" without using calculators. The use of calculators was *not* allowed for the grade 4 Mathematics tests.

CHAPTER 5
DESIGN OF THE SCIENCE & TECHNOLOGY ASSESSMENT

BACKGROUND

The Science & Technology section of the MCAS is based on the learning standards described in the Massachusetts *Science & Technology Curriculum Framework* (1996). These learning standards were developed in collaboration with teachers, school and district administrators, scientists, technology experts, college faculty, parents, and representatives of business and community organizations across the state.

The MCAS Science & Technology tests were designed to assess two fundamental dimensions of learning: content knowledge and skills in using and applying science and technology.

CONTENT STRANDS

Four major content strands identified by the *Science & Technology Curriculum Framework* serve as the foundation for the MCAS Science & Technology tests and its reporting categories:

- Inquiry
- Domains of science:
  - Physical sciences
  - Life sciences
  - Earth and space sciences
- Technology
- Science, technology, and human affairs

Table 5-2 shows the approximate distribution of MCAS Science & Technology items by content strand and substrand for each grade level. For reporting purposes, MCAS questions were linked with the reporting category that most closely represents the standard(s) assessed.

Table 5-2
Approximate Distribution of MCAS Science & Technology Test Questions
By Content Strand and Substrand

Content Strand	Substrands	Grade 4	Grade 8	Grade 10
–	In accordance with the *Science & Technology Curriculum Framework* and assessment design, many questions that address other content strands will also be inquiry-based, and are therefore not limited to a specific percentage of questions.			
Domains of Science	Physical Sciences	25%	25%	25%
	Life Sciences	25%	25%	25%
	Earth and Space Sciences	25%	25%	25%
Technology	The Design Process	5%		5%
	Understanding and Using Technology	15%	15%	15%
Science, Technology, and Human Affairs		5%	5%	5%

SKILLS IN USING AND APPLYING SCIENCE & TECHNOLOGY

In addition to content knowledge, students were expected to demonstrate various process skills fundamental to science and technology. Critical investigation and problem-solving skills included:

- observation;
- hypothesis formulation and testing; and
- evaluation and use of evidence to propose, design, and test solutions.

For the purposes of the MCAS Science & Technology tests, these scientific and technology-related process skills were grouped into three major areas: thinking skills, procedural skills, and application skills.

Thinking Skills

Items in this area assessed student understanding of concepts. In order to demonstrate thinking skills, students were required, for example, to recognize, evaluate, analyze, and explain natural scientific and technological phenomena.

Procedural Skills

Items in this area assessed student knowledge and understanding of scientific and technological procedures.

Application Skills

Items in this area assessed student skill in selecting appropriate scientific and technological concepts and procedures and appropriately applying these concepts and procedures to solve real-life and theoretical problems.

TYPES OF SCIENCE & TECHNOLOGY QUESTIONS ON MCAS

Two types of questions were used at each grade level:

- multiple-choice; and
- open-response.

Students were required to answer questions that assessed the content knowledge and process skills that are developmentally appropriate for each grade level.

Table 5-2 presents the approximate number of items for each item type for each component in each grade.

Table 5-2
Approximate Distribution of Science & Technology Items by Type

Grade	Item Type	Number of Test Items (per student test booklet)
4 and 8	Multiple-choice	32
	Open response	7
10	Multiple-choice	38
	Open response	10

Table 5-3 describes the exact number of items that appeared in the 1998 MCAS Science & Technology tests. Note that technology and science, technology, and human affairs were collapsed and referred to as technology.

Table 5-3
Distribution of Items, 1998 Science & Technology Test

Grade	Reporting Category	Common		Matrix (Total Across 12 Forms)	
		Multiple-Choice	Open Response	Multiple-Choice	Open Response
4	Inquiry	5	2	6	3
	Physical Sciences	4	0	15	3
	Life Sciences	5	1	17	2
	Earth & Space Sciences	6	1	17	2
	Technology	6	2	17	2
	Total	26	6	72	12
8	Inquiry	3	1	12	0
	Physical Sciences	5	2	16	3
	Life Sciences	6	1	14	3
	Earth & Space Sciences	5	1	14	2
	Technology	7	1	16	4
	Total	26	6	72	12
10	Inquiry	1	1	9	0
	Physical Sciences	8	0	24	7
	Life Sciences	6	4	22	5
	Earth & Space Sciences	7	2	20	4
	Technology	10	1	21	8
	Total	32	8	96	24

CHAPTER 6
TEST DEVELOPMENT PROCESS

As described in the preceding chapters, MCAS tests were developed to meet a complex set of content and cognitive specifications. In addition, to provide accurate measurement across four performance categories, MCAS items need to demonstrate acceptable statistical characteristics. To ensure an adequate selection of items to build final test forms, twice as many items were developed as were ultimately needed.

MCAS tests have been designed and developed by the Massachusetts Development of Education in collaboration with committees of Massachusetts educators (Assessment Development Committee) and the Department's testing contractor. Assessment Development Committees for the areas of English Language Arts, Mathematics, and Science & Technology have met regularly since January 1996 to develop test blueprints and specifications, and test items and scoring guides based on the *Massachusetts Curriculum Framework* learning standards in these content areas. In addition to the Assessment Development Committees, the Department convened a Bias Review Committee to review individual test items and accompanying materials and to recommend editing or removal of items that were likely to place a particular group of students at an advantage or disadvantage for non-educational reasons. Table 6-1 presents the major steps in the MCAS test development process. Additional information about the process follows the table.

Table 6-1
Major Steps in the MCAS Test Development Process

	Step	When Occurred
1	Assessment Development Committee (ADC) test blueprint development	January 1996
2	Item writing	April–June 1996
3	Internal item review	July–August 1996
4	Assessment Development Committee item review	August 1996
5	Item editing	September–December 1996
6	Item tryout form assembly	March 1997
7	Item tryout review	April 1998
8	Item tryout administration	April 28–May 9, 1997
9	Item tryout scoring	May–June 1997
10	Item tryout data analysis	July 1997
11	Initial item selection	September–October 1997
12	Assessment Development Committee selection and editing of common and matrix items	December 1997
13	DOE-contractor review	January 1998
14	External bias and sensitivity review	March 1998
15	DOE-contractor bias and sensitivity resolution	March 1998
16	Operational test assembly	February–March 1998
17	Edit drafts of operational tests	March 1998
18	Braille translation	March 1998
19	Spanish translation	March 1998

At the early meetings of the Assessment Development Committees, test specifications and designs were reviewed and item ideas were generated. Item ideas ranged from broad-brush, "addition of two two-digit numbers with renaming (carrying) in a story problem" to targeted, "addition of two-digit numbers with renaming in a story problem that asks about the number of pieces of equipment in a park" to writing a complete draft item. The contractor's test developers expanded upon the item ideas and edited the items

BEST COPY AVAILABLE

for technical accuracy and adherence to sound testing practice. Refined items were later presented to the Assessment Development Committees for review and revision.

INTERNAL ITEM REVIEW

- Lead or peer test developer within the content specialty reviewed the typed item, open-response scoring guide, and any reading selections and graphics.
- The content reviewer considered item content and structure; appropriateness to designated content area; item format; clarity; ambiguity; developmental appropriateness and quality of items; reading selections and graphics; appropriateness of scoring guide descriptions and distinctions; and, for multiple choice items, the presence of a single correct answer.
- The content reviewer also considered whether the scoring guide adequately addressed the possible range of performance on the item.
- Fundamental questions for the content reviewer to ask included, but were not be limited to, the following:
  –What is the item asking?

  –Is the key the only possible key?

  –Is the open-response item scorable as written (correct words used to elicit response defined by guide)?
  –Is the wording of the scoring guide appropriate and parallel to the item wording?
  –Is the item complete (e.g., with scoring guide, content codes, key, grade level, and contract identified)?
  –Is the item appropriate for the designated grade level?

ITEM EDITING

Editors reviewed and edited the items from the ADC item review to ensure uniform style (based on *The Chicago Manual of Style, 14th Edition*) and adherence to sound testing principles. These principles stipulated that items:

- were correct with regard to grammar, punctuation, usage, and spelling;
- were written in a clear, concise style;
- were unambiguous in explaining to students what is expected for a maximum score;
- were written at a reading level that allowed students to demo nstrate his or her knowledge of the tested subject matter;
- exhibited high technical quality regarding psychometric characteristics;
- had appropriate answer options or score-point descriptors; and
- were free of bias and sensitivity concerns.

ITEM TRYOUT FORM ASSEMBLY

Multiple test forms were created for English language arts, mathematics, and science and technology for each grade level (4, 8, and 10). Within each form, test questions were grouped by content (e.g., in order to form a more homogeneous criterion for item analysis, tryout forms were not built to be parallel). See section on Operational Test Assembly for more details of this process.

ITEM TRYOUTS

Following initial test development, a tryout of questions in Mathematics and Science & Technology was administered to all students in grades 4, 8 and 10 in the spring of 1997. A tryout of English Language Arts questions was administered in the fall of 1997. No student, school, district or state results were reported for any tryout. Item statistics generated by the item tryouts were used to review, revise, and make final selections of questions for the MCAS tests administered in 1998.

The tryouts were designed to mirror the administration of the operational assessment program. The tryout test forms were spiraled so that each school would have some students taking each test form and each test form would be administered to a random sample of students. All public school students in grades 4, 8, and 10 in all schools in Massachusetts were required to participate in the tryout.

ITEM TRYOUT SCORING

Responses to multiple-choice items were optically scanned. Responses to open-response items were scored using a consensus-scoring model, that is, rather than developing a training pack with benchmark papers, a group of highly experienced scorers used scoring rubrics to guide discussion of student responses and come to mutually acceptable scores.

ITEM TRYOUT DATA ANALYSIS

The following statistics were calculated for each multiple-choice item: item difficulty (percent correct), item discrimination (point-biserial correlations), item quartile distribution (distribution of student responses or scores within each quartile of the criterion score distribution), and differential item functioning (DIF) statistics comparing males and females and white and black student responses.

These statistics were calculated for short-answer questions, except there were insufficient students to calculate DIF statistics for white-black comparisons. Statistics calculated for open-response items were identical to those calculated for short-answer questions, except the Pearson product-moment correlation was used rather than the point-biserial correlation.

INITIAL ITEM SELECTION

Test developers selected acceptable items to present to the Assessment Development Committees based on statistical information (see Table 6-2 for the format in which information was provided), comments from scorers and their own professional judgement regarding the quality of items. Note, not all item statistics were computed for item tryout items.

Table 6-2
Format of Item Statistics

Sample:	*A*			*Score Point*		*n*	*% of Total*	*% of 1st quartile*	*% of 2nd quartile*	*% of 3rd quartile*	*% of 4th quartile*	*Mean crit. score*
				OR	*MC*							
Criterion	*B*			*BL*	*BL*	*R*	*S*	*T*	*U*	*V*	*W*	*X*
				0	*A*							
Difficulty (Mn): C		*Discrimination (r): D*		*1*	*B*							
A: E	*c: F*	*b(01): G*	*b(12): H*	*2*	*C*							
Fit: K		*b(23): I*	*b(34): J*	*3*	*D*							
I(s12): L	*I(s23): M*		*I(s34): N*	*4*	*E*							
DIF(F-M): O	*DIF(B-W): P*		*DIF(H-W): Q*	*T*	*T*	*Y*						*Z*

BEST COPY AVAILABLE

A A description of the sample is entered here, such as: "1999 Massachusetts grade 4 item tryout sample for mathematics."

B The criterion measure used for biserial correlations and differential item functioning analyses is entered here, such as: "Form 12 Total Mathematics score."

C Classical item difficulty or item mean. For multiple-choice items this is equivalent to percent of students responding correctly (p-value); for open-response items this is equivalent to the average student item score.

D Classical item discrimination statistic. For multiple-choice items this is a corrected point-biserial correlation; for open-response items, this is a Pearson product-moment correlation (a corrected item-to-total score correlation).

E Item response theory item discrimination parameter.

F Item response theory lower asymptote (guessing) parameter (for the three-parameter logistic model). Used only for multiple-choice or other items where student guessing might lead to a correct answer.

G Item response theory difficulty parameter for differentiating scores of 0 and 1. There is one difficulty parameter for multiple-choice items, and one between each pair of consecutive score categories for open-response items.

H Item response theory difficulty parameter for differentiating scores of 1 and 2. This will be blank for multiple-choice items.

I Item response theory difficulty parameter for differentiating scores of 2 and 3. This will be blank for multiple-choice items.

J Item response theory difficulty parameter for differentiating scores of 3 and 4. This will be blank for multiple-choice items.

K Item response theory fit statistic, describing how well the IRT model fits the item's data.

L Amount of information item provides for differentiating between students at the first and second client-set performance standards. Requires that performance standards are already set. The sum of item information at these performance standard cut-points is directly related to the test's decision accuracy.

M Amount of information item provides for differentiating between students at the second and third client-set performance standards. Requires that performance standards are already set.

N Amount of information item provides for differentiating between students at the third and fourth client-set performance standards. Requires that performance standards are already set.

O Standardized difference between matched (by weighting to total group on criterion score) samples of male and female students. Significance of difference based on Mantel-Haenszel statistic and indicated by one asterisk (.01 level) or two asterisks (.001 level).

P Standardized difference between matched (by weighting to total group on criterion score) samples of white and black students.

Q Standardized difference between matched (by weighting to total group on criterion score) samples of white and Hispanic students.

R For open-response or multiple-choice items, the number of examinees who left this question blank. For open-response, the next five rows present the number of students with scores of 0, 1, 2, 3, and 4 respectively. More rows are added if there are additional score points. For multiple-choice items, those rows indicate the number of examinees who chose options A, B, C, D, and E, respectively.

S For each row in this column, the percent of examinees with each score (open-response) or who chose each option (multiple-choice) is indicated.

T Of those examinees scoring in the top quartile on the total criterion score, the percent whose response was blank. The next five rows present similar information for the other score points.

U Of those examinees scoring in the second quartile on the total criterion score, the percent whose response was blank. The next five rows present similar information for the other score points.

V Of those examinees scoring in the third quartile on the total criterion score, the percent whose response was blank. The next five rows present similar information for the other score points.

W Of those examinees scoring in the lowest quartile on the total criterion score, the percent whose response was blank.

X Mean total criterion score of those examinees whose score point was blank. For following rows, the mean criterion score is given for examinees achieving other score points. For multiple-choice items, this should be highest for the correct option. For open-response items, the means should be ordered for score points 0 to 4, and spread reasonably well.

Y Total sample size.

Z Sample mean on the criterion.

EXTERNAL BIAS AND SENSITIVITY REVIEW

A bias and sensitivity review committee of educators was convened to review items and English Language Arts reading passages for potential bias and sensitivity issues. Bias is defined as question context or content that is irrelevant to the curriculum being assessed that affects test scores of an identifiable subgroup of students. Sensitivity refers to issues that are not related to the curriculum being assessed and might offend or distract students. Items that received comment during the bias and sensitivity review were reviewed at a meeting between senior Department staff and the contractor to consider the Bias Review Committee's recommendations and make final decisions for item selection.

SELECTION OF COMMON AND MATRIX ITEMS

Test developers presented item statistics to the Assessment Development Committees to assist in the Committees' recommendations for placement of items into the common and matrix portions of the test. The final decision for selections was made by the Department of Education with the assistance of the testing contractor.

OPERATIONAL TEST ASSEMBLY

Test assembly is the sorting and laying out of item sets into test forms. Criteria considered during this process included the following:

- Content coverage/match to test design. The curriculum specialist completed an initial sort of items into sets based on a balance of content categories across sessions and forms, as well as a match to the test design (number of multiple-choice, short-answer, and open-response items).
- Item difficulty and complexity. Item statistics resulting from data analysis of previously tested items were used to assure similar levels of difficulty and complexity across forms.
- Visual balance. Item sets were reviewed to ensure that each reflected a similar length and "density" of selected items (e.g., length/complexity of reading selections, number of graphics).
- Option balance. Each item set was checked to verify that it contains a roughly equivalent number of key options (As, Bs, Cs, and Ds).
- Name balance. Item sets were reviewed to ensure diversity of names used.
- Bias. Each item set was reviewed to ensure fairness and balance based on gender, ethnicity, religion, socio-economic status, and other factors.

- Page fit. Item placement was modified to ensure the best fit and arrangement of items on any given page.
- Facing page issues. For multiple items that are associated with a single stimulus (graphic or reading selection), consideration was given to whether the group needs to begin on a left- or right-hand page, as well as to the nature and amount of material that needed to be on facing pages. These considerations serve to minimize the amount of "page flipping" required of the students.
- Relationships between forms. The set of "common" items must be placed identically in each version of the forms. Matrix-sampled item sets differ from form to form, but must take up the same number of pages in each form so that sessions and content areas begin on the same page in every form. Therefore, the number of pages needed for the longest form often drives the layout of each form.
- Visual appeal. The visual accessibility of each page of the form is always considered, including such aspects as the amount of "white space," the density of the text, and the number of graphics.

EDIT DRAFTS OF OPERATIONAL TESTS

Any changes that the test construction specialist makes are reviewed and approved by the test developer. Once a form is laid out in what is considered its final form, the form is read through to identify any final considerations, including the following:

- Editorial changes. All text is scrutinized for editorial accuracy, including consistency of instructional language, grammar, spelling, punctuation, and layout. The contractor's publishing standards are based on *The Chicago Manual of Style, 14th Edition*.
- "Keying" items. Items are reviewed for any information that may "key" (or provide information that would help answer) another item. Decisions about moving keying items are based on the severity of the key-in and the placement of the items in relation to each other within the form.
- Key patterns. The final sequence of keys is reviewed to ensure that their order appears random (e.g., no recognizable pattern, no more than three of the same key in a row).

BRAILLE AND LARGE PRINT TESTS

One form of each of the May 1998 MCAS tests was translated into Braille by a subcontractor specializing in test materials for blind and visually-handicapped students. Additionally, one form of each of the May 1998 MCAS tests was adapted into a large print version.

SPANISH TRANSLATION

One form of the May 1998 MCAS mathematics and science and technology tests were adapted into Spanish. The Spanish version of the MCAS tests were presented in a bilingual format (Spanish/English) with identical test items presented on opposing pages: left-facing pages presented items in Spanish; right-facing pages presented identical items in English. This format was adopted based on field testing a Spanish only adaptation and a bilingual format adaptation among Limited English Proficient (LEP) students in approximately 10 public school districts.

In adapting a test to another language, a number of decisions have to be made. Depending on the nature of the original test, on the target language, and the intended examinee population, the adapted test may be very similar or quite different from the original. In this case, because intended examinees were known to come from different Hispanic countries, representing a variety of dialects rather than a single dialect, it was decided to use standard Spanish in the test, and to include certain dialectal variants as a gloss in brackets as needed. Because of the nature of the subjects being tested (math and science), and their link to the state standards, it was agreed ahead of time that the basic content of the tests should remain the same if possible.

BEST COPY AVAILABLE

There were a number of steps in the adaptation of MCAS for Spanish-speaking students. A preliminary review of the instruments showed that only two items needed to be replaced with items from other test forms in English. The two items identified in the review involved assumed knowledge of American culture. For example, one item assumed knowledge of how American football is played.

Another change that was made in the instruments involved translating English names to Spanish (James = Jaime), provided the names were easily translatable.

Two native speakers of Spanish were identified. Each was a professional translator with knowledge of item writing procedures and experience in test translation and test translation review. Each translator was a specialist in either math or science. The translator of the mathematics test had an undergraduate degree in mathematics from a university in Paraguay. The science translator had a degree in medical anthropology from a university in Colombia. Both had experience translating standardized tests, and had previously received instruction on item writing.

Both translators were oriented to the project. The orientation included information on the MCAS program and the most frequent countries of origin of examinees who would take the MCAS in Spanish. Subsequently, the translators began work on the first draft. Their first draft was reviewed by a senior translation specialist, who made initial decisions about how to handle wording common to both tests, such as that found in the instructions, headers, footers, item stems, etc. The senior translation specialist then sent each translator's work to the other with instructions that the translation be evaluated by comparing it line by line and item by item with the English version. The comments of each reviewer were reviewed, and then forwarded to the original translator with further observations or recommendations.

The DOE collected systematic feedback from teachers and students on the Spanish version following its administration. The feedback elicited from teachers concerning Spanish usage in the math and science tests showed that they felt the Spanish version accurately reflected the English original.

SECTION II
TEST ADMINISTRATION

CHAPTER 7
TEST ADMINISTRATION

RESPONSIBILITY FOR ADMINISTRATION

As indicated in the *Principal's Administration Manual* (Massachusetts Department of Education, 1998e), principals were responsible for the proper administration of the MCAS. Directors of charter schools, 766-approved private schools, institutional school programs, and educational collaboratives were responsible for the compliance with administration requirements in their school. Manuals and certification forms were used to ensure uniformity of administration procedures across schools.

PROCEDURES

Principals were instructed to read the *Principal's Administration Manual* thoroughly prior to testing and to be familiar with the instructions given in the *Test Administrator's Manual* (Massachusetts Department of Education, 1998f). The chapter "Conducting Test Administration" in the *Test Administrator's Manual* contains sections that detail the procedures that were to be followed for each test session. The chapter also contains the actual scripts "to be read aloud to students AS PRINTED during test administration" (p. 9). Another critical document produced and disseminated by the Department of Education was *The Massachusetts Comprehensive Assessment System: Requirements for Test Scheduling, Student Participation, and Test Security and Ethics* (Massachusetts Department of Education, 1998g).

ADMINISTRATOR TRAINING

In addition to the two administration manuals, the Massachusetts Department of Education made a training videotape available to all schools in early April 1998. Eight additional broadcasts of the training were carried on cable television.

TEST ADMINISTRATION SCHEDULE

MCAS testing materials were received in schools the week of April 27, 1998. The test administration window was from May 4 through May 22, 1998. The Department of Education supplied schools with sample test administration schedules for grades 4, 8, and 10. Table 7-1 presents the grade 10 sample test administration schedule.

Table 7-1
1998 Grade 10 Sample Test Administration Schedule

- Seventeen 45-minute test sessions, plus one 20–30 minute session for completion of student identification information, questionnaire, and an optional practice test
- Two 45-minute sessions per day maximum
- Makeup sessions scheduled throughout the three weeks as necessary

May 1998				
Monday	Tuesday	Wednesday	Thursday	Friday
4 Student Identification Questionnaire and Practice Test (30 min.)	5 English Language Arts English Language Arts	6 English Language Arts English Language Arts	7 English Language Arts	6 English Language Arts English Language Arts
11	12 Mathematics	13 Mathematics Mathematics	14 Mathematics	15 Science & Technology Science & Technology
18	19 Science & Technology Science & Technology	20 History and Social Science Item Tryout History and Social Science Item Tryout	21	22

PARTICIPATION REQUIREMENTS

All public school students in grades 4, 8, and 10 were required to participate in the MCAS, per the Educational Reform Act of 1993, including students enrolled in charter schools, and students receiving publicly funded special education in 766-approved private schools, institutional schools, and collaboratives.

Students with Disabilities

Students with disabilities were defined as students with an Individualized Education Plan (IEP) or a plan of instructional accommodations provided under Section 504 of the Rehabilitation Act of 1973. For such students, the IEP plan of the Section 504 team is required to consider the following questions in determining how a student will participate:

- Can this student take the tests under routine conditions?
- If the student is not able to take the tests under routine conditions, will he or she be able to take these tests if appropriate test accommodations are provided?
- If a student cannot take the tests, even with accommodations, what would be an appropriate alternative assessment to enable the student to demonstrate his or her knowledge of the standards contained in the curriculum frameworks?

Limited English Proficient Students

Limited English Proficient (LEP) students were defined as students who met any of the following conditions:

- were enrolled in a Transitional Bilingual Program;
- received English as a Second Language support;
- were not born in the United States and whose native language was a language other than English and who were currently not able to perform ordinary classroom work in English; or
- were born in the United States to non-English speaking parents and who were not currently able to perform ordinary classroom work in English.

LEP students were required to participate in the MCAS if they met either of the following criteria:

- student had been enrolled in school in the United States for more than three years; or
- student was in a Transitional Bilingual Education program or received English as a Second Language support and had been/would be recommended for regular education classes for the 1989–99 school year.

Requirements for Spanish-Speaking LEP Students

Spanish-speaking LEP students who have completed three or more years of school in the United States were required to take the English language version of MCAS.

Spanish-speaking LEP students who do not yet have the fluency to participate in the English language version of the MCAS were required to participate in the Spanish language version of the mathematics and science and technology tests if they met all of the following criteria:

- had completed three or fewer years of school in the United States;
- were in a Transitional Bilingual Education program or received English as a Second Language support and were not to be recommended for regular education classes for the 1989–99 school year; and
- possessed reading and writing skills in Spanish appropriate to their grade level.

Accommodations

The Massachusetts Department of Education published a list of appropriate accommodations in *The Massachusetts Comprehensive Assessment System: Requirements for Test Scheduling, Student Participation, and Test Security and Ethics* (Massachusetts Department of Education, 1998g).

TEST SECURITY

Strict question and test security measures were implemented during all phases of development and production in order to maintain the fairness and integrity of the MCAS. To this end, each of the MCAS administration manuals contains a chapter on "Test Security and Ethics." In the chapter, it is stated

> *The quality and usefulness of the assessment data generated by MCAS depends, in large part, on uniformity of test administration and security of test materials. Valuable information about student achievement and curriculum effectiveness will be seriously compromised if test security is not strictly implemented and maintained (p. 5).*

The chapter includes sections on penalties, school/principal's responsibilities, and instructions to be given to students regarding the use of test materials. The school/principal's responsibilities include

- taking inventory of testing materials received by the school,
- monitoring the distribution and use of these materials, and
- ensuring the complete and error-free return of all materials.

ACCOUNTING FOR TEST MATERIALS

The administration manuals also contained explicit instructions for the handling of test booklets, answer documents, and other materials. Material tracking and verification forms were provided to principals and test administrators to help them account for test materials. Upon completion of testing, test administrators assembled the test materials for return to the principal. Used response documents were separated from unused ones and were packaged in special envelopes provided to schools. The school principal organized the testing materials, using the material verification form, to verify the return of all secure testing materials to the testing contractor.

Each principal received detailed instructions and a prepaid, pre-printed air-bill for returning test materials to the testing contractor. Principals were instructed to call the shipping contractor toll free when their materials were ready for pickup after testing. Shipped packages were completely and easily traceable. Personnel were able to track a particular package any time from date of pickup to date of delivery. A toll-free number was also provided to principals to provide notification of any problems or delays with pickup.

The outside of each box containing test materials was labeled by school and district. Upon receipt of each box, the labels were checked and the boxes were logged in. The resulting list was compared to a master distribution file on a daily basis. One week after the close of the testing window, a list of outstanding schools or missing boxes was produced, and applicable schools were contacted for discrepancy resolution.

Once boxes were scanned, they were placed on a holding skid (by grade) to be processed. In order to ensure accuracy, each person who checked materials worked with only one school at a time.

During log-in, staff opened boxes and reviewed administration forms. If any of the administration forms were missing, the school was contacted. A log-in supervisor used the principal's certification forms to enter into an electronic spreadsheet the following information:

- the number of materials sent to the school,
- the number of materials returned from the school, and
- the date the materials were logged into the spreadsheet.

In addition, the following information was entered into the spreadsheet and updated:

- the name of individual who logged in the materials,
- whether or not the school had a discrepancy and the date any discrepancy was sent to the school for resolution, and
- whether the school or the Department of Education has resolved the discrepancy.

The newly created spreadsheet was then compared to the master distribution file to determine if any discrepancies existed. If there was a difference between the number of materials sent to the school and the number received from the school, the discrepancy resolution process began.

Once the materials were accounted for, all demographic sheets were removed from the response booklets and placed under a school header pre-slugged with school name, school code, and the number of students in that school. This became the official file upon which school reports were based.

The used response booklets were processed by hand to check their general condition and to remove any unnecessary materials. Schools with materials that were returned with significant problems were reported to the school and the Department of Education. Efforts were made to correct gridding problems, and any missing or damaged headers were replaced.

About two percent of the total test forms were received from the schools in poor condition and could not be scanned. Unscannable forms were manually entered into the system. Large-print response booklets were also entered manually.

After the booklets were checked, they were oriented in one direction and boxed by school. The school header sheet was placed on the top of booklets in the box, which was then sent for scanning.

SECTION III
DEVELOPMENT AND REPORTING OF SCORES

CHAPTER 8
SCORING

Student answer booklets were scanned so that all information necessary to score responses and produce reports was captured and converted into an electronic format. This conversion included all student identification and demographic information, school information, multiple-choice data, and digital image clips of hand-written responses. This chapter summarizes the score processing procedures for the MCAS.

Student responses to multiple-choice questions were machine scored. Responses to all other questions were read and evaluated individually by trained readers.

MACHINE-SCORED ITEMS

Student responses to multiple-choice were optically scanned. The scoring key was applied to the captured item responses. Correct answers were assigned a score of one point; incorrect answers were assigned a score of zero points. Multiple-choice questions were used within all subject area tests: English Language Arts, Mathematics, and Science & Technology.

ITEMS SCORED BY READERS

Digital imaging and a computerized scoring system were used in the scoring process for all short-answer, open-response questions and short compositions. Digital imaging allowed electronic copies of students' responses for a single item to be sent to readers who scored the responses. The computerized scoring system assigned student responses to readers. It provided maximum randomization of student work, to ensure that no one reader, or small group of readers, scored multiple responses from the same school. It also provided continuous monitoring of the performance of readers, allowing leadership staff to rescore student responses and retrain readers when necessary. Scoring methods for each type of open-response question are described in the following three subsections.

SCORING GUIDES FOR SHORT-ANSWER ITEMS

Short-answer questions, used on the Mathematics test, were hand-scored by contractor scoring staff. Correct answers were assigned a score of one point; incorrect answers were assigned a score of zero points based on an item-specific scoring guide. Most short-answer questions had a single correct numeric answer. In some cases, there were multiple acceptable answers (see Figure 8-1) or a range of correct answers (for example, correct answer: a number in the range of 356 to 358). One grade 10 short-answer question was somewhat more complex to score (correct answer: any set of 9 numbers with a range of 20, mean of 85, and median of 85; e.g., 75, 75, 75, 80, 85, 90, 95, 95). Figure 8-1 presents an example of a short-answer item with its scoring guide.

Figure 8-1
Example of a Short-Answer Item and Its Scoring Guide

Item	Write a RULE to find the next number in the pattern. 90, 87, 84, 81, ___
Scoring guide	Score as correct: Subtract 3 -3 minus 3

SCORING GUIDES FOR OPEN-RESPONSE ITEMS

Item-specific scoring guides were developed by test development staff for each open-response item prior to scoring. Figure 8-2 presents an example of a scoring guide for an open-response item.

SCORING GUIDE FOR WRITING PROMPTS

Each students was required to write one long and one short composition in response to writing prompts. Each composition was assigned a score for Topic/Idea Development (on a 1–6 scale) and a score for Standard English Conventions (on a 1–4 scale). Readers for the long and short compositions included contractor scorers and teachers at three Massachusetts Writing Institutes. The *MCAS Writing Scoring Guide* in Figure 8-3 was used for scoring all compositions. In addition to the scores, "analytic annotations" (scorer comments) were also used in reporting. These are comments on topic development, organization, details, language/style, sentences, grammar and usage, and mechanics, as shown in Figure 8-3.

Figure 8-2
Example of an Open-Response Item and Its Scoring Guide

Item	To make a house handicapped accessible, a ramp is being constructed to the floor of the porch. The Americans with Disabilities Act requires that a ramp have an incline of no more than 5. Assume that the maximum allowable angle is used and that the floor of the porch to which the ramp is constructed is 4 feet above the ground. (You may refer to the trigonometric table on your Mathematics Reference Sheet.) a. Draw and label a picture showing the ramp and porch. b. Based on the information above, how far is the end of the ramp from the porch? Show your work. c. Based on the information above, what is the length of the ramp? Show your work.
Scoring guide	Score 4 if The student scores 5 points Score 3 if The student scores 4 points Score 2 if The student scores 3 or 2 points Score 1 if The student scores 1 point Score 0 if Response is totally incorrect or irrelevant. Score Blank if No response Scoring information: Part a: 1 point for correct drawing of porch and ramp For drawing, the student must show right triangle with angle of 5 and 4' for length of vertical leg of right triangle opposite the 5 angle. Part b: 1 point for correct distance from porch = 45.71 feet 1 point for correct strategy displayed through work, e.g., $\tan 5^{\circ} = 0{,}0875 = 4/x$ $x = 4/0{,}0875 = 45.71$ feet Note: Other correct approaches are acceptable.) Part c: 1 point for correct length of ramp = 45.9 feet 1 point for correct strategy displayed through work, e.g., $45.71^2 + 4^2$ = length of ramp ^{2} $(2089.4 + 16)^{.5}$ = length of ramp = 45.9 feet OR

	$\sin 5 = 4/r$ $r = 4/\sin 5$ $r = 45.9$ feet (or 45.87; 45.89) Some numbers in work may vary due to rounding, but answers should be correct to at least the nearest tenth of a foot. If rounding is to nearest foot, work must show ramp longer than horizontal distance before rounding. Note: If student reverses order of b and c, credit can be awarded as above, provided work/diagram shows student understands which length he/she found.

MCAS WRITING SCORING GUIDE (LONG COMPOSITION)

Topic/Idea Development

1	2	3	4	5	6
• Little topic/idea development, organization, and/or details • Little or no awareness of audience and/or task	• Limited or weak topic/idea development, organization, and/or details • Limited awareness of audience and/or task	• Rudimentary topic/idea development and/or organization • Basic supporting details • Simplistic language	• Moderate topic/idea development and organization • Adequate, relevant details • Some variety in language	• Full topic/idea development • Logical organization • Strong details • Appropriate use of language	• Rich topic/idea development • Careful and/or subtle organization • Effective/rich use of language

Analytic Annotations

		Commendations		Needs	
Topic/Idea Development	The overall effect of the paper	TX TY	Effective development of topic or ideas Original development of topic or ideas	TJ TK	More development of the topic Better understanding of the writing task
Organization	The degree to which the response is • focused • clearly and logically ordered • clarified by paragraphs	OX OY	Evidence of planning Consistently focused from beginning to end	OJ OK	Organization of ideas Transitions between ideas
Details	The degree to which the response includes examples that develop the main points	DX DY	Carefully chosen, relevant details Develops details that support the topic	DJ DK	More effective choice of details Development of details
Language/ Style	The degree to which manipulation of language, including vocabulary, word choice, word combination, and sentence variety is effectively achieved	LX LY	Word choice enhances meaning Language creates distinctive tone or style	LJ LK	More variety/richness in word choice More variety in sentence structure

Standard English Conventions

1	2	3	4
• Errors seriously interfere with communication AND • Little control of sentence structure, grammar and usage, and mechanics	• Errors interfere somewhat with communication and/or • Too many errors relative to the length of the essay or complexity of sentence structure, grammar and usage, and mechanics	• Errors do not interfere with communication and/or • Few errors relative to length of essay or complexity of sentence structure, grammar and usage, and mechanics	• Control of sentence structure, grammar and usage, and mechanics (length and complexity of essay provide opportunity for student to show control of standard English conventions)

Analytic Annotations

		Commendations		Needs	
Structure	The degree to which the response includes sentences that are correct in structure	SP	Correct sentence structure	SR	Correct sentence structure
Grammar and Usage	The degree to which the response demonstrates correct • use of standard grammatical rules of English • word usage and vocabulary	GUP GUQ	Correct application of grammatical rules Control of vocabulary and word usage	GUR GUS	Correct application of grammatical rules Greater attention to correct word usage
Mechanics	The degree to which the response demonstrates correct • spelling • capitalization • punctuation	MP MQ	Control of mechanics aids clarity Correct mechanics in sophisticated construction	MR MS	Greater control of mechanics More careful proofreading

BEST COPY AVAILABLE

SELECTION OF SCORING STAFF

Scoring was led by a scoring director and scoring site managers who managed the various scoring locations. Chief readers, curriculum specialists, were responsible for managing the technical aspects of scoring including hiring quality assurance coordinators, overseeing the development of training materials, and ensuring training is implemented properly.

Chief readers worked with quality assurance coordinators and human resource specialists to hire qualified readers. For the scoring of MCAS, readers were required to have completed two years of college, but were preferred to have earned a four-year college degree. In addition, readers were required to have an appropriate background for the discipline they scored. Applicant screening procedures included

- a formal, structured interview;
- reference checks; and
- a review of each returning reader's documented history on scoring projects similar to MCAS to ensure that the contractor is not hiring a reader who has not demonstrated successful work as a reader.

Table 8-4 summarizes the qualifications of the 1998 MCAS readers.

Table 8-4
Qualifications of 1998 MCAS Scorers

Scoring Responsibility		Educational Credentials				Teaching Experience	Total
		Doctorate	Masters	Bachelors	Other		
Leadership	n	5	30	17	1	38	53
	%	9%	57%	32%	2%	71%	100%
Readers	n	235		326	240	373	801
	%	29%		41%	30%	47%	100%

There are three additional points to be made about scoring staff qualifications.

- Data in Table 8-4 do not include approximately 720 Massachusetts educators who scored a portion of the writing assessments as part of Department of Education-sponsored writing institutes;
    - teaching experience ranged from one to thirty-two years; and
    - among the readers, information collected about advanced degrees did not differentiate doctoral degrees from masters degrees.

READER TRAINING AND QUALIFICATION

For each item, quality assurance coordinators explained how the anchor pack papers exemplified the descriptors of the score points. After discussion of the anchor pack, readers attempted to score a training pack containing exemplars correctly. The quality assurance coordinators then reviewed the training pack and answered any questions readers had before actual scoring began. Subsequently, quality assurance coordinators monitored the scoring process and provided further training on any given item as warranted. Readers were required to maintain an acceptable scoring accuracy rate.

SCORING PROCESS

For short-answer and open-response questions, scoring was controlled by an electronic image scoring management system, which distributed digital images of student responses to readers. These responses were randomly assigned to readers. Thus, the probability is low that any reader would score more than one item from a particular student's response booklet. By using the maximum possible number of readers for each student, this procedure effectively minimized error variance due to reader sampling.

All readers had at their workstations a complete set of scoring materials (i.e., scoring guides, training packs) for each of the items. Quality assurance coordinators were available to advise and assist readers with their scoring efforts.

Quality assurance coordinators or other highly experienced scorers (verifiers) performed a series of *read-behinds* in which they scored responses previously scored by readers. Quality assurance coordinators used the agreement rates from these read-behinds to provide ongoing feedback to the readers.

For each question, about 10% of the responses were rescored as a read-behind and about 1% of the responses were scored independently by two readers using a double blind process.

Monitoring Scoring

The scoring management system tracked reader accuracy throughout the scoring process. After a reader scored a student response, the management system determined whether that response should also be scored by another reader, scored by a quality assurance coordinator or other scoring official, or routed for special attention[1]. Quality assurance coordinators and other scoring officials could get current reader accuracy reports and speed reports on-line at any time. Summary or detailed reports could be produced for any time period. Such capability served to ensure reliable and valid scoring.

The weighted averages of and total (exact or adjacent) percent agreement are reported in Table 8-5. Exact agreement is defined as both readers assigned the paper the same score, and adjacent agreement is defined as the two readers scores differed by one point. The weighting was based on the number of responses that were rescored for each question. Note, these data may underestimate scorer accuracy. Blanks were included

1 Student responses indicating possible child abuse or suicidal tendencies were flagged by readers for school attention.

in both the read-behind and double-blind rescoring. Readers were instructed to score as zero any question for which the student had made a mark of any kind. But in many instances it was impossible for the reader to tell whether there was a mark on the page written by the student or whether there was a crease in the paper, bleed-through from the other side of the page or dust on the image screen. In such instances, these responses were counted as neither exact nor adjacent agreement, though the effect of blanks and zeroes on student scores was identical.

Table 8-5 1998 MCAS Scoring Agreement Rates on Open-Response and Short-Answer Questions						
Grade	Reading		Mathematics		Science & Technology	
	Read behind	Double Blind	Read behind	Double Blind	Read behind	Double Blind
4	99.1	94.9	99.5	99.0	99.3	96.9
8	99.0	95.5	99.0	98.3	99.5	97.7
10	99.2	97.5	98.9	97.2	99.2	97.6
Agreement rates include exact agreement, in which two readers assigned the same score to a student response, and adjacent agreement, in which the scores assigned by two readers differed by no more than one point.						

WRITING PROMPTS

Two readers independently scored all long compositions. If the two scores were not in exact or adjacent agreement, the two readers discussed and re-evaluated the composition to reach agreement on a score. By this method, the process of correcting inaccurate scores served as a way to prevent reader drift and provide continuous training. The final score for the long compositions was the sum of the scores assigned by the two readers.

Only one reader scored each short composition. Short compositions were responses to matrix prompts; thus scores on short compositions were not used in the computation of scaled scores or performance levels. Samples of the scores assigned by readers to both short and long compositions were regularly verified using the read-behind and double-blind methods to ensure the quality of the scores.

CHAPTER 9
STANDARD SETTING

PERFORMANCE LEVEL NAMES AND GENERAL DESCRIPTIONS

Standard setting is the process of determining the minimum, or threshold score, for each performance level, grade, and subject area for which results are reported. The multistep process of setting standards for the MCAS tests of May 1998 began in February 1998, when the Massachusetts Board of Education adopted general descriptions for each of the four performance levels to be used in reporting. These general descriptions were the basis for all standard-setting activities.

- *Advanced:* Students at this level demonstrate a comprehensive and in-depth understanding of rigorous subject matter, and provide sophisticated solutions to complex problems.
- *Proficient:* Students at this level demonstrate a solid understanding of challenging subject matter and solve a wide variety of problems.
- *Needs Improvement:* Students at this level demonstrate a partial understanding of subject matter and solve some simple problems.
- *Failing:* Students at this level demonstrate a minimal understanding of subject matter and do not solve even simple problems.

SUBJECT-SPECIFIC PERFORMANCE LEVEL DEFINITIONS

Building on the general definitions, content specialists developed general performance level definitions for each subject area. These definitions were further refined for each grade level. Those descriptions were approved by the Board in June 1998 and were used in the standard-setting process.

In August 1998, the Department of Education convened panels of Massachusetts educators and non-educators to participate in the standard-setting process for MCAS. This process resulted in the identification of a minimum total test score (threshold score) for each performance level, by grade and subject area.

It is important to recognize that standard setting is not the same as scoring, which is the process of assigning score points to student responses. Scoring must occur **before** standard setting can begin. MCAS scoring took place from June through August 1998, and the standard setting-process began in August.

PANELISTS

Twelve panels were convened to set performance standards for the MCAS—one panel for each grade level (4, 8, and 10) in four areas—1) language and literature (reading), 2) composition (writing), 3) mathematics, and 4) science and technology. Two hundred and nine (209) panelists participated in two full days of meetings to set the performance level standards. The panels were composed of educators, parents and business leaders, and members of the general public. Table 9-1 presents data regarding the background of the panelists.

Table 9-1
Background of Standard-Setting Panelists

Background	Number	Percent

Classroom Teachers	106	51
Administrators	45	22
Higher Education	15	7
Business Community	35	17
School Committees or Local/State Government	8	3
Total	209	100

PROCESS

The panelists used the Body of Work (BoW) standard-setting method. The hallmark of the BoW method is that panelists examine complete student response sets (student responses to multiple-choice questions and actual student work on open-response questions) and match each student response set to one of the MCAS performance level categories. This is done in three major steps: 1) training/calibration, 2) range finding, and 3) pinpointing.

Training/Calibration

During this first phase of the MCAS standard-setting process, panelists reviewed all MCAS test questions for their assigned content area and grade level, and content- and grade-specific descriptors for each performance level. Panelists were given the opportunity to discuss and comment on test questions and descriptors. Next, to ensure that panelists attained a common interpretation of performance descriptors and the relationship of those descriptors to student work, panel members individually assigned performance levels to a set of six sample student responses. Panelists then compared their individual results and discussed at length how the performance level descriptors supported their conclusions.

Range-Finding

During the range-finding phase of standard setting, identical sets of student work that spanned the score continuum were provided to each panelist. Panelists were asked to independently categorize the sets as *Advanced*, *Proficient*, *Needs Improvement*, or *Failing*, based on the performance level descriptors. This process revealed which sets of student work generated the most agreement and which generated the most disagreement among panelists. The results were documented, and the sets of work that generated the most disagreement defined the score intervals in which the threshold scores must fall.

Pinpointing

Additional sets of student work from score ranges that generated disagreement were presented to panelists. Panelists assigned performance levels to these sets of responses. The minimum score for each performance level was precisely pinpointed by determining the score around which there was, collectively, the maximum disagreement between panelists. This is the point that best represents the transition from response sets at a higher level to those at a lower level.

Following is a detailed description of the steps followed in implementing the MCAS standard-setting design.

Before the Meeting

1. For each subject-grade combination (e.g., grade 8 mathematics) pinpointing folders were prepared from samples of student work. This sample was double-scored to increase the accuracy of the standard-setting process. Any students

whose body of work was of uneven quality (for example, some open-response questions with scores of four and others with scores of one) were excluded, as were students whose open-response and multiple-choice responses were particularly discrepant. Folders ranged in scores from the highest obtained score in the remaining sample to the "approximately chance level" (.25 times the number of multiple-choice items plus one times the number of open-response items). Each folder consisted of five sets of student work at each of four score points (e.g., five 12s, five 13s, five 14s, and five 15s), with the exception of the top folder (folder with highest scores). The top folder differed because there often were fewer than five papers available at any particular score point. Thus, the twenty papers in the top folder covered a wider range of scores. Approximately ten pinpointing folders were created for each subject-grade combination.

2. Range-finding folders were prepared from the pinpointing folders. The highest-scoring and two lowest-scoring papers were selected from each pinpointing folder. Thus, range- finding folders had about thirty samples of student work.
3. For each subject-grade combination, six student response sets spanning the range of performance were identified from the pinpointing folders. The facilitator reviewed the sets and prepared training notes consisting of points to be made during discussion of those student response sets. Focus was on ways that student responses illustrate characteristics described in the performance level definitions.
4. The Massachusetts Department of Education created a list of members of each panel (one panel per subject area, four subject areas per grade, and three grades), ensuring each group had the proper diversity of membership (educator, parent, policy-maker, businessperson, ethnicity, gender, etc.). Color-coded name tags were provided to panel members.

General Meeting

1. Before the panels broke into separate groups, there was a general session at which logistical issues were addressed and the standard-setting procedures

BEST COPY AVAILABLE

explained by the chief of standard setting. Major steps of the panel meeting portion of the meeting were described.

Panel Meeting

1. Facilitators distributed the descriptor of a four-point response to each open-response question. Panel members were asked to review and discuss the test questions—open-response and multiple-choice. (Panelists had been asked to answer the questions before the meeting, and they were to have brought with them the tests and the performance level definitions. Additional copies were distributed to those who needed them.)
2. The facilitators led a discussion of the performance level definitions.
3. Training folders were distributed to every judge. The multiple-choice display at the end of a set were pointed out. Facilitators explained that it too should be considered when judgments are being made about the student work.
4. Judges were asked to rank independently the six previously identified student response sets based on overall quality, keeping in mind the performance level descriptions. Each judge listed the six student serial numbers in rank order from high to low performance on a separate piece of paper.
5. While the judges rank ordered the six student response sets, the facilitator wrote the serial numbers of the six sets on an overhead transparency in a vertical list in order from highest performance to lowest performance. When the judges completed their rankings, the facilitators showed the score rankings on the overhead projector and had the judges note the extent of agreement.
6. Judges were asked to assign each of the six response sets to a performance level. They each wrote the performance level initials (A, P, N, or F) next to the student serial numbers they listed in rank order in step 4.
7. Facilitators drew four columns to the right of the six serial numbers on the overhead transparency, and labeled the columns A, P, N, and F. Facilitators recorded the judges' ratings (based on shows of hands) next to the serial numbers on the overhead.
8. Facilitators lead a discussion of the six response sets as they related to the performance levels.
9. The heterogeneous range-finding folders were distributed to every judge. The facilitators pointed out the multiple-choice display at the end of a set, and explained that it too should be considered when judgments are being made about the student work.
10. Facilitators distributed a Range-Finding Rating Form to each judge, and asked the judges to enter their names in the name boxes and encode a home telephone number in the "ID" field. Judges were given the opportunity to reconsider their ratings of the six student response sets and transfer their "final" ratings to the Range-Finding Rating Form on which the serial numbers for these and other response sets in the range-finding folder had been entered in order from high to low performance.
11. Judges were asked to decide independently the performance levels of the rest of the student response sets in the range-finding folder and record their ratings on their Range-Finding Rating Forms in the left set of columns.
12. Judges' ratings were recorded on the "Range-Finding" overhead transparency, based on shows of hands. Judges were asked to view the overhead and decide if they wanted to change their minds regarding any of the student response sets. Group discussion was allowed. Changed ratings were recorded in the "Second Ratings" columns of the Range-Finding Rating Form.
13. When the judges completed step 12, their materials were collected. From these data, the chief of standard setting determined the pinpointing folder or folders that must be evaluated by the judges for determining each of the three cut points.
14. For each pinpointing folder, the performance level decision to be made was indicated, e.g.,

    Folders 3 and 4—*Advanced* or *Proficient*?
    Folders 9 and 10—*Proficient* or *Needs Improvement*?
    Folder 15—*Needs Improvement* or *Failing*?

BEST COPY AVAILABLE

15. The group of judges was divided into three small groups. Each small group examined the folder or folders for one cut score[2]. Each judge independently completed a Pinpointing Rating Form, including the name boxes and ID field, for each folder he or she was assigned. Materials were rotated so all three small groups examined the folder or folders for every cut point.
16. All standard-setting materials (ranking sheets, forms, folders, tests, definitions, etc.) were collected and returned to the chief of standard setting.

As panelists turned in their materials, they were given an evaluation form to fill out and were invited to return at 4:30 to see a summary of the results.

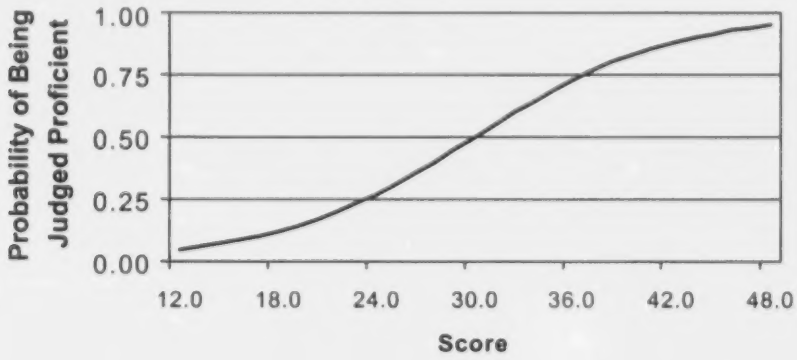
Panelists' Evaluation of Process

On a 1 to 5 scale with 5 being most positive, the average panelist ratings were 4.5 regarding clarity of instructions, 4.8 regarding level of understanding, and 4.3 regarding confidence in ratings.

Data Analysis

Data were analyzed using logistic regression. A separate logistic regression was run for each threshold decision. The unit of analysis was a panelist's decision about a single student's body of work. Test scores were used to predict the probability of a student's work being classified as meeting or exceeding each performance level. Figure 9-2 provides a graphical example of the results of a logistic regression.

Figure 9-2
Graphical Example of Logistic Regression Results
Note, in Figure 9-2, it is at a test score of 30 that the probability of being judged proficient is .5. Thus, 30



would be the minimum score at which a student would be considered *Proficient*.

Results

Reading and Writing (Composition) threshold determinations were based on independent panels. The final threshold determination for English Language Arts was based on the sum of the threshold

[2] The purpose of dividing the group into thirds was to reduce the need for multiple copies of folders. This way, each group worked with one-third of the folders, finished the work on one cut score, and then passed the folders to the next group for them to do the same.

BEST COPY AVAILABLE

recommendations of the two component parts. Table 9-3 presents the final threshold determinations that were presented to the Massachusetts Board of Education and approved at their September 1998 meeting.

Table 9-3
Threshold (Minimum) Total Test Score For Each Performance Category

Grade	Subject Area	Maximum Score on Test	Threshold Score		
			Advanced	Proficient	Needs Improvement
4	English Language Arts	68	59.37	46.46	23.74
	Mathematics	50	39.88	31.70	18.21
	Science & Technology	50	39.45	29.81	18.07
8	English Language Arts	68	57.71	41.00	27.16
	Mathematics	50	42.68	32.50	22.48
	Science & Technology	50	39.66	29.52	22.14
10	English Language Arts	84	66.83	51.49	36.95
	Mathematics	60	45.63	34.39	23.80
	Science & Technology	62	46.68	34.61	21.72

Standard Errors of Estimate for Threshold Scores

Table 9-4 presents the standard errors of estimate for the results of the logistic regressions. Standard errors were estimated by applying the logistic regression technique separately to each panelist's data. Thus, for each threshold decision, there was a distribution of estimated thresholds. The standard error was estimated as the standard deviation of that distribution divided by the square root of the number of panelists. Standard errors were estimated separately for Reading and Writing.

Table 9-4
Standard Errors of Logistic Regressions For Each Performance Category

Grade	Subject Area	Maximum Score on Test	Standard Error		
			Advanced	Proficient	Needs Improvement
4	Reading	48	.22	.56	.45
	Writing	20	.22	.31	.36
	Mathematics	50	.33	.24	.80
	Science & Technology	50	.28	.52	.53
8	Reading	48	.33	.63	.34
	Writing	20	.27	.28	.20
	Mathematics	50	.46	.61	.46
	Science & Technology	50	.21	.39	.51
10	Reading	64	.56	.42	.50
	Writing	20	.27	.16	.08
	Mathematics	60	.45	.58	.55
	Science & Technology	62	.80	.59	.72

CHAPTER 10
SCALING

The MCAS tests were designed to measure student performance against the learning standards described in the *Curriculum Frameworks*. Consistent with this purpose, primary results on the MCAS tests are reported in terms of performance levels that describe student performance in relation to these established state standards. There are four performance levels: *Advanced*, *Proficient*, *Needs Improvement*, and *Failing*, as described in Chapter 9. Students received a separate performance level classification (based on scaled score) for each test. School and district performance level results were reported as the number and percentage of students who attained each performance level at each grade level tested.

In addition to performance levels, MCAS results are reported as scaled scores. Scaled scores in each content area range from 200 to 280. Scaled scores supplement the MCAS performance level results by providing information about the position of a student's results within a performance level. School- and district-level scaled scores are calculated by computing the average of student-level scaled scores.

TRANSLATING RAW SCORES TO SCALED SCORES (SCALING)

Students' raw scores, or total number of points, on the MCAS tests are translated to scaled scores using a process called scaling. Scaling simply converts raw points from one scale to another. Converting from raw scores to scaled scores does not change the rank ordering of students, give more weight to particular questions, or change students' performance level classifications.

Linear scaling parameters were determined so the minimum scaled score for *Needs Improvement* was 220, the minimum scaled score for *Proficient* was 240, and the minimum scaled score for *Advanced* was 260 for each MCAS test. This was done by solving two linear equations relating the raw threshold scores to these predetermined scaled score values. The resulting functions that translate raw scores to scaled scores are:

$$S = m_1 r + b_1 \quad \text{if } r < P, \text{ and}$$
$$S = m_2 r + b_2 \quad \text{if } r > P$$

where S is the scaled score, r is the raw score, and P is the *Proficient* threshold. The values of the ms and the bs are shown in Table 10-1.

Table 10-1
Transformation Constants Used to Compute Scaled Scores

Grade	Subject Area	Transformation Constants			
		m_1	b_1	m_2	b_2
4	English Language Arts	0.88	198.10	1.55	167.00
	Mathematics	1.48	192.10	2.44	161.55
	Science & Technology	1.70	188.23	2.07	177.15
8	English Language Arts	1.45	179.76	1.20	189.95
	Mathematics	2.00	174.09	1.96	175.17
	Science & Technology	2.71	158.95	1.97	180.76
10	English Language Arts	1.38	168.15	1.30	171.89
	Mathematics	1.89	174.01	1.78	177.85
	Science & Technology	1.55	185.30	1.65	181.63

After the transformation constants were applied, scores were rounded to the nearest even integer. Transformed scores below 200 were reported as 200; transformed scores above 280 were reported as 280.

In any given year, test form difficulty and rounding might lead to some scaled scores between 200 and 280 not being attainable. For the 1998 MCAS, for all subjects and grades 200 was an obtainable value. Table 10-2 reports the highest and lowest attainable scaled scores on the 1998 MCAS.

Table 10-2
Minimum and Maximum Obtainable Scores on the 1998 MCAS

Grade	Subject Area	Raw Score		Scaled Score	
		Minimum	Maximum	Minimum	Maximum
4	English Language Arts	0	68	200	272
	Mathematics	0	50	200	280
	Science & Technology	0	50	200	280
8	English Language Arts	0	68	200	272
	Mathematics	0	50	200	274
	Science & Technology	0	50	200	280
10	English Language Arts	0	84	200	280
	Mathematics	0	60	200	280
	Science & Technology	0	62	200	280

CHAPTER 11
SCORE REPORTING

Table 11-1 lists the primary MCAS reports.

Table 11-1 Primary MCAS Reports	
1.	*Student Report for Parents/Guardians*
2.	*Student Labels*
3.	*School Test Item Analysis Report*
4.	*District Test Item Analysis Report*
5.	*School Report*
6.	*District Report*
7.	*Union Report*
8.	*1998 Statewide Summary of District Performance on the Massachusetts Comprehensive Assessment System (MCAS)*
9.	*MCAS Student Results CD*
10.	*MCAS School and District Results CD*
11.	*Report of 1998 Statewide Results: The Massachusetts Comprehensive Assessment System (MCAS)*

STUDENT REPORT FOR PARENTS/GUARDIANS

Student reports show the scaled score for each subject area, as well as a score band that indicates the standard error of measurement surrounding each score. General performance level definitions are provided so that parents/guardians will understand how to interpret the scaled scores. Information is also provided about how the student performed in each subject subarea, compared to his/her overall performance in the subject area[1]. Specific comments are provided about the student's writing performance. Information is also provided to show how the student's performance compared to the average scores from the student's school, district, and state. An overview of test content is provided, along with a cautionary statement about interpreting scores and guidelines for parents/guardians for helping their children improve. The report also indicates that the child's school should be contacted if there are any questions about the child's report.

The Department of Education provides additional documentation, *Understanding Your MCAS 1998 Student Report for Parents/Guardians* (Massachusetts Department of Education, 1998h), which explains in detail how to interpret student reports. This interpretive manual is available in English, Cape Verdean, Chinese, Haitian, Kmer, Portuguese, Russian, Spanish, and Vietnamese. In addition, although all student reports were printed in English, report shells were available in the aforementioned languages to aid parents and guardians in interpreting their child's report.

STUDENT LABELS

To aid schools in keeping track of student scores, schools were supplied with student score information on individual labels that they could affix to student files, if desired.

---

[1] This information proved to be somewhat difficult to interpret and will be removed from this report in future years. Other options for reporting student performance in subject subareas will be explored.

BEST COPY AVAILABLE

SCHOOL AND DISTRICT TEST ITEM ANALYSIS REPORT

The *Test Item Analysis Report* shows the answers that each student gave on the common multiple-choice questions, as well as his/her score the common writing prompt and on each common open-response question. The report also summarizes overall performance at the school, district, and state levels for each of the question types.

Each school receives a separate *Test Item Analysis Report* for each subject area and grade. The report is designed to be used in conjunction with the publication *The Massachusetts Comprehensive Assessment System: Release of May 1998 Test Items* (Massachusetts Department of Education, 1998a), which contains all common test questions. When the report and the publication are used together, educators are provided with a detailed picture of student performance. The *Guide to Interpreting the 1998 MCAS School and District Reports* (Massachusetts Department of Education, 1998i) also explains the *Test Item Analysis Report* in detail.

SCHOOL, DISTRICT, AND UNION REPORTS

The school, district, and union reports are intended for administrators and other interested parties. The school report includes performance level definitions, scaled score intervals, student status definitions, and information about how summary statistics are affected by students not tested; all of which are intended to help the reader interpret the report. The school report provides all results for the school, the district, and the entire state. The results provided are

- the number of students tested by student status (regular, students with disabilities, and limited English proficient students) for all subject areas combined and separately for each subject area,
- the percentage of students in each performance level by subject area,
- the distribution of scaled scores by subject area,
- the number of students in each performance level by subject area and student status,
- subscores by subject subarea and by question type,
- three-year comparisons of school results, and
- average subject score by number of years in the school or district.

The district report is the same as the school report, except that it does not include the school-level data and the three-year comparisons are by district rather than by school. The *Guide to Interpreting the 1998 MCAS School and District Reports* (Massachusetts Department of Education, 1998i) explains the school and district reports in detail.

The union report is analogous to the district report, but is prepared for school unions—sets of districts sharing a single superintendent.

1998 STATEWIDE SUMMARY OF DISTRICT PERFORMANCE ON THE MASSACHUSETTS COMPREHENSIVE ASSESSMENT SYSTEM (MCAS)

The *1998 Statewide Summary of District Performance on the Massachusetts Comprehensive Assessment System (MCAS)* (Massachusetts Department of Education, 1998j) summarizes performance of all districts in the state, providing a page of information for each.

MCAS STUDENT RESULTS CD

The student results CD is an electronic version of the *Test Item Analysis Report*. Districts were provided with a CD containing confidential student data for each school in the district.

MCAS SCHOOL AND DISTRICT RESULTS CD

The *MCAS School and District Results CD* is an electronic version of the *1998 Statewide Summary of District Performance on the Massachusetts Comprehensive Assessment System (MCAS)*.

REPORT OF 1998 STATEWIDE RESULTS: THE MASSACHUSETTS COMPREHENSIVE ASSESSMENT SYSTEM (MCAS)

The *Report of 1998 Statewide Results: The Massachusetts Comprehensive Assessment System (MCAS)* (Massachusetts Department of Education, 1998k) presented statewide participation rates, performance levels, and scaled score results.

CHAPTER 12
STATE RESULTS

This chapter presents key participation and performance results from the May 1998 MCAS administration.

Table 12-1
Students Tested[1] on the MCAS Tests of May 1998

Grade Level	Enrolled	Percent Tested in English Language Arts	Percent Tested in Mathematics	Percent Tested in Science & Technology	Tested in all Content Areas	
					Number	Percent
4	76,365	97.4	98.4	98.4	74,382	97.4
8	70,053	97.0	97.7	97.7	67,991	97.1
10	62,462	95.1	95.9	95.9	59,376	95.1
Total	208,880	96.6	97.4	97.4	201,749	96.6

[1] Includes regular education students, students with disabilities, and limited English proficient students.

Table 12-2
Regular Students Tested on the MCAS Tests of May 1998

Grade Level	Enrolled	Percent Tested in English Language Arts	Percent Tested in Mathematics	Percent Tested in Science & Technology	Tested in all Content Areas	
					Number	Percent
4	60,977	99.6	99.8	99.8	60,807	99.7
8	57,603	99.0	99.3	99.3	57,143	99.2
10	52,371	97.5	97.7	97.7	51,096	97.6
Total	170,951	98.8	99.0	99.0	169,064	98.9

Table 12-3
Students with Disabilities Tested on the MCAS Tests of May 1998

Grade Level	Enrolled	Percent Tested in English Language Arts	Percent Tested in Mathematics	Percent Tested in Science & Technology	Tested in all Content Areas	
					Number	Percent
4	12,497	94.1	95.2	95.2	11,705	93.7
8	10,844	93.6	94.3	94.0	10,084	93.0
10	8,286	91.9	92.5	92.5	7,562	91.3
Total	31,627	93.4	94.2	94.1	29,351	92.8

Table 12-4
Limited English Proficient Students Tested[1] on the MCAS Tests of May 1998

Grade Level	Enrolled	Percent Tested in English Language Arts	Percent Tested in Mathematics	Percent Tested in Science & Technology	Tested in all Content Areas[2]	
					Number	Percent
4	2,891	66.0	82.8	82.7	1,870	64.7
8	1,606	47.2	64.9	65.0	764	47.6
10	1,805	39.7	58.2	58.4	718	39.8
Total	6,302	53.7	71.2	71.2	3,352	53.2

[1] Spanish-speaking limited English proficient students who had been in school in the United States for three or fewer years (as of May 1998) for whom the English version of MCAS was not appropriate were required to participate in the Spanish version of MCAS. The difference in percentages of students participating across all three subject areas is largely due to the fact that the Spanish version of MCAS included tests in Mathematics and Science & Technology **only**.

In grades 4, 8, and 10, there were 509, 270, and 154 students, respectively, who were identified by school personnel as both students with disabilities and students with limited English proficiency. These students are not included in this table; these students are included in Table 12-3: Students with Disabilities Tested on the MCAS Tests of May 1998.

[2] Only limited English proficient students who were in school in the United States for more than three years (as of May 1998) were required to participate in the English version of MCAS, which included tests in all three content areas.

Table 12-5
1998 Statewide MCAS Performance Level Results by Student Status
Grade 4
(percentage of students at each performance level)[1]

Content Area	Student Status	Scaled Scores	Performance Level				
			Advanced	Proficient	Needs Improvement	Failing (Tested)	Failing (Absent)[2]
English Language Arts	All	230	1	19	66	15	0
	Regular	233	1	22	69	8	0
	S w/ Disabilities	221	0	3	54	43	0
	LEP	219	0	2	47	51	0
Mathematics	All	234	11	23	44	23	0
	Regular	236	13	26	44	17	0
	S w/ Disabilities	223	2	10	42	46	0
	LEP	217	2	5	28	65	0
Science & Technology	All	238	6	42	40	12	0
	Regular	240	7	48	38	7	0
	S w/ Disabilities	228	1	22	50	27	0
	LEP	221	1	9	41	49	0

S w/ Disabilities - Students with Disabilities; LEP – Limited English Proficient

1 Percentages may not total 100 percent due to rounding.

2 For the purpose of computing school, district, and state results, students who were absent from any subject area MCAS test were assigned the minimum scaled score of 200 and a performance level of Failing for that subject area.

Table 12-6
1998 Statewide MCAS Performance Level Results by Student Status
Grade 8
(percentage of students at each performance level)[1]

Content Area	Student Status	Scaled Scores	Performance Level				
			Advanced	Proficient	Needs Improvement	Failing (Tested)	Failing (Absent)[2]
English Language Arts	All	237	3	52	31	13	1
	Regular	240	3	60	29	7	1
	S w/ Disabilities	222	0	15	41	44	1
	LEP	219	0	13	34	52	1
Mathematics	All	227	8	23	26	41	1
	Regular	230	10	27	28	34	1
	S w/ Disabilities	210	1	5	15	78	1
	LEP	209	1	6	13	79	1
Science & Technology	All	225	2	26	31	40	1
	Regular	228	2	30	34	33	1
	S w/ Disabilities	211	0	6	18	75	1
	LEP	207	0	3	10	86	1

S w/ Disabilities – Students with Disabilities; LEP – Limited English Proficient

1 Percentages may not total 100 percent due to rounding.

2 For the purpose of computing school, district, and state results, students who were absent from any subject area MCAS test were assigned the minimum scaled score of 200 and a performance level of Failing for that subject area.

Table 12-7
1998 Statewide MCAS Performance Level Results by Student Status
Grade 10
(percentage of students at each performance level)[1]

Content Area	Student Status	Scaled Scores	Performance Level				
			Advanced	Proficient	Needs Improvement	Failing (Tested)	Failing (Absent)[2]
English Language Arts	All	230	5	33	34	26	2
	Regular	233	6	38	35	20	2
	S w/ Disabilities	213	0	7	27	64	3
	LEP	214	0	8	28	59	5
Mathematics	All	222	7	17	24	50	2
	Regular	225	8	19	27	44	2
	S w/ Disabilities	206	1	3	9	84	4
	LEP	208	1	5	12	78	4
Science & Technology	All	225	1	21	42	34	2
	Regular	227	2	24	45	28	2
	S w/ Disabilities	213	0	4	25	67	4
	LEP	211	0	2	19	75	4

S w/ Disabilities – Students with Disabilities; LEP – Limited English Proficient

1 Percentages may not total 100 percent due to rounding.

2 For the purpose of computing school, district, and state results, students who were absent from any subject area MCAS test were assigned the minimum scaled score of 200 and a performance level of Failing for that subject area.

SECTION IV
TECHNICAL
CHARACTERISTICS

CHAPTER 13
ITEM ANALYSES

As noted in Brown (1983), "a test is only as good as the items it contains." A complete evaluation of a test's quality must include an evaluation of each question. Both the *Standards for Educational and Psychological Testing* and the *Code of Fair Testing Practices in Education* include standards for identifying quality questions. Questions should assess only knowledge or skills that are under assessment and should avoid assessing irrelevant factors. They should also be unambiguous and free of grammatical errors, potentially insensitive content or language, and other confounding characteristics. Further, questions must not unfairly disadvantage test takers from particular racial, ethnic, or gender groups.

Both qualitative and quantitative analyses are conducted to ensure that MCAS questions meet these standards. Previous sections in this report have delineated the qualitative checks on question quality. The current chapter focuses on more quantitative evaluations. The statistical evaluations are presented in three sections: 1) difficulty indices, 2) item-test correlations, and 3) subgroup differences in item performance. The results presented in this chapter are based on the statewide administration of MCAS in May of 1998. About 75,000 grade 4 students, 68,000 grade 8 students, and 58,000 grade 10 students participated in the assessment.

DIFFICULTY INDICES

All multiple-choice, short-answer, and open-response questions were evaluated in terms of difficulty and relationship to overall score according to standard classical test theory practice. Difficulty was measured by averaging the proportion of points received across all students who received the question. Multiple-choice and short-answer questions were scored dichotomously (correct v. incorrect), so for these questions, the difficulty index is simply the proportion of students who correctly answered the question. Open-response questions allowed for scores between 0 and 4. By computing the difficulty index as the average proportion of points received, the indices for multiple-choice, short-answer, and open-response questions are placed on a similar scale; the index ranges from 0 to 1 regardless of the question type. Although this index is traditionally described as a measure of difficulty (as it is described here), it is properly interpreted as an "easiness index" because larger values indicate easier questions. An index of 0 indicates that no student received credit for the question, and an index of 1 indicates that every student received full credit for the question.

ITEM-TEST CORRELATIONS

Within classical test theory, these relationships are assessed using correlation coefficients that are typically described as either item-test correlations or, more commonly, discrimination indices. The discrimination index used to analyze MCAS multiple-choice items and short-answer items, which are scored 0 or 1, was the point-biserial correlation between item score and a criterion total score on the test. For open-response items, item discrimination indices were based on the Pearson product-moment correlation. The theoretical range of these statistics is from –1 to 1, with a typical range from .3 to .6.

Discrimination indices can be thought of as measures of how closely a question assesses the same knowledge and skills assessed by other questions contributing to the criterion total score. That is, the discrimination index can be interpreted as a measure of construct consistency. In light of this interpretation, the selection of an appropriate criterion total score is crucial to the interpretation of the discrimination index. For MCAS, appropriate criterion scores were selected based on item type and function (common or matrix). The selected criterion scores are provided in Table 13-1. For example, the criterion score for common open-response and short-answer items was the total score on all common multiple-choice, open-response, and short-answer items.

BEST COPY AVAILABLE

Table 13-1 Criterion Score Used in Computing the Discrimination Index For Each Item Type and Function					
Item Type	Item Function	Scores Included in the Total			
		MC Common	MC Matrix	OR & SA Common	OR & SA Matrix
Multiple-Choice (MC)	Common	✓			
	Matrix	✓	✓		
Open Response (OR) and Short Answer (SA)	Common	✓		✓	
	Matrix	✓	✓	✓	✓
Writing Prompt (WP)	Common	✓		✓	
	Matrix	✓	✓	✓	✓

For the writing prompt, the reading score was used as the criterion.

SUMMARY OF ITEM ANALYSIS RESULTS

Frequency distributions and summary statistics of the difficulty and discrimination indices for each question are provided in Appendix B and summarized in Table 13-2. Both Appendix B and Table 13-2 also provide separate distribution information for common and matrix multiple-choice questions.

Table 13-2 Average Difficulty and Discrimination of Different Question Types For Each Subject and Grade											
Grade	Questions		Reading			Mathematics			Science & Technology		
			n	Diff	Disc	*n*	Diff	Disc	*n*	Diff	Disc
4	MC	All	124	0.61	0.36	81	0.61	0.34	98	0.64	0.32
		Common	28	0.61	0.38	21	0.61	0.35	26	0.65	0.32
		Matrix	96	0.61	0.36	60	0.62	0.33	72	0.64	0.32
	Short Answer		-	-	-	17	0.5	0.37	-	-	-
	Open Response		29	0.44	0.49	18	0.47	0.55	18	0.46	0.43
8	MC	All	124	0.66	0.37	81	0.54	0.35	98	0.6	0.32
		Common	28	0.68	0.34	21	0.58	0.36	26	0.57	0.29
		Matrix	96	0.66	0.37	60	0.53	0.35	72	0.62	0.33
	Short Answer		-	-	-	17	0.52	0.49	-	-	-
	Open Response		29	0.47	0.56	18	0.38	0.64	19	0.37	0.54
10	MC	All	128	0.64	0.35	111	0.45	0.32	128	0.56	0.3
		Common	32	0.66	0.34	27	0.55	0.37	32	0.58	0.29
		Matrix	96	0.63	0.35	84	0.42	0.3	96	0.55	0.3
	Short Answer		-	-	-	17	0.41	0.46	-	-	-
	Open Response		32	0.43	0.59	32	0.24	0.62	32	0.22	0.52

SUBGROUP DIFFERENCES IN QUESTION PERFORMANCE

The *Code of Fair Testing Practices in Education* explicitly states that subgroup differences in performance should be examined when sample sizes permit, and actions should be taken to make certain that differences in performance are due to construct-relevant, rather than irrelevant, factors. The *Standards for Educational and Psychological Testing* includes similar guidelines. As part of the effort to identify such problems, MCAS questions were evaluated in terms of differential item functioning (DIF) statistics.

DIF procedures are designed to identify questions for which subgroups of interest perform differently beyond the impact of differences in overall achievement. For MCAS, the standardization DIF procedure (Dorans and Kulick, 1986) was employed to evaluate two subgroup pairs: male v. female and white v. black[1]. This procedure calculates the difference in item performance for groups of students matched for achievement on the total test. That is, the average item performance is calculated for students at every total score, then an overall average is calculated weighting the total score distribution so it is the same for the two groups.

The index ranges from –1 to 1 for multiple-choice and short-answer questions and is adjusted to the same scale (by dividing by four) for open-response questions. Negative numbers indicate that the question was more difficult for female or black students. Positive numbers indicate that the question was easier for female or black students.

Dorans and Holland (1993) suggested that index values between –0.05 and 0.05 should be considered negligible for dichotomously scored questions (such as MCAS multiple-choice and short-answer questions). Most MCAS multiple-choice and short-answer questions fall within this range. Dorans and Holland further stated that dichotomously scored questions with values between –0.10 and –0.05 and between 0.05 and 0.10 (i.e., "low" DIF) should be inspected to ensure that no possible effect is overlooked, and that questions with values outside the [–0.10, 0.10] range (i.e., "high" DIF) are more unusual and should be examined very carefully. These standards can be applied to open-response questions by accounting for the larger range of possible index values and scaling appropriately. That is, values of the DIF index can range from –4.0 to 4.0, so the corresponding ranges are between –0.2 and 0.2 for negligible difference, between –0.4 and –0.2 and between 0.2 and 0.4 for "low" DIF and outside [-0.4, 0.4] for "high" DIF.

DIF indices indicate differential performance between two groups. That differential performance may or may not be indicative of bias in the test. Course-taking patterns, group differences in interests, or differences in school curricula can lead to DIF. If subgroup differences in performance are related to construct-relevant factors, the questions should be considered for inclusion on a test.

Each question was categorized according to the guidelines adapted from Dorans and Holland (1993). Tables 13-3 and 13-4 provide the number of questions in each of the three DIF categories for male-female and white-black comparisons.

Table 13-3
Number of Questions in Each Male-Female DIF Category:

Grade	DIF Level	English Language Arts		Mathematics			Science & Technology	
		MC	OR	MC	SA	OR	MC	OR
4	Negligible	100	26	75	16	17	77	18
	Low	21	3	6	1	1	21	0
	High	3	0	0	0	0	0	2
8	Negligible	106	25	71	17	13	69	14
	Low	15	4	9	0	5	26	1
	High	3	0	1	0	0	3	0
10	Negligible	113	30	92	15	30	92	29
	Low	14	2	16	2	1	32	4
	High	1	0	3	0	1	4	0

[1] The Mantel-Haentzel procedure was also used to determine DIF during the test development process. Items with statistically significant DIF were flagged and indicated in the statistical information presented to the Bias and Sensitivity Review Committee.

Table 13-4 Number of Questions in Each White-Black DIF Category								
Grade	DIF Level	English Language Arts		Mathematics			Science & Technology	
		MC	OR	MC	SA	OR	MC	OR
4	Negligible	109	26	66	11	16	98	18
	Low	13	3	13	6	2	0	0
	High	2	0	2	0	0	0	0
8	Negligible	92	29	64	15	16	77	16
	Low	28	0	17	2	4	19	2
	High	4	0	0	0	0	2	0
10	Negligible	107	31	90	12	30	104	32
	Low	16	1	20	4	2	17	1
	High	5	0	1	1	0	5	0

CHAPTER 14
RELIABILITY

Although an individual test question's performance is an important focus for evaluation, a complete evaluation of an assessment must also address the way that questions function together and complement one another. Any measurement includes some amount of measurement error; that is, no measurement can be perfectly accurate. This is true of academic assessments—no assessment can measure students perfectly accurately; some students will receive scores that underestimate their true ability, and other students will receive scores that overestimate their true ability. Questions that function well together produce assessments that have less measurement error; that is, the errors made should be small on average. Such assessments are described as reliable.

There are a number of ways to estimate an assessment's reliability. One approach is to split all test questions into two groups and then correlate students' scores on the two half tests. This is known as a split-half estimate of reliability. If the two half-test scores correlate highly, questions on the two half tests must be measuring very similar knowledge or skills. This is evidence that the questions complement one another and function well as a group. This also suggests that measurement error will be minimal.

The split-half method requires the psychometrician to select which questions contribute to each half-test score. This decision may have an impact on the resulting correlation. Cronbach (1951) provided a statistic that avoids this concern about the split-half method: Coefficient Alpha (a).

RELIABILITY AND STANDARD ERRORS OF MEASUREMENT

Table 14-1 presents descriptive statistics, Cronbach's a coefficient, and raw and scaled score standard errors of measurement for each subject area (English Language Arts, Mathematics, and Science & Technology), separately for each grade level. The item analysis sample excludes students who did not take one or more sections of the subject.

Note, two scaled-score standard errors of measurement are presented: one for scaled scores below 240 and one for scaled scores of 240 and above. This is because different slopes are used in the linear transformation to scaled scores at these two different parts of the scaled score range.

Table 14-1
Reliabilities, Standard Errors of Measurement and Descriptive Statistics

Grade	Subject	n	Raw Score						Scaled Score	
									<240	>=240
			Min.	Max.	Mean	S.D.	Rel.	S.E.M.	S.E.M.	S.E.M.
4	English Language Arts	73,527	4	67	36.4	10.9	0.90	3.5	3.0	5.4
	Mathematics	74,068	0	50	26.8	9.9	0.87	3.6	5.4	8.9
	Science & Technology	74,069	0	49	28.5	8.0	0.86	3.0	5.1	6.3
8	English Language Arts	66,707	4	67	40.9	10.4	0.90	3.3	4.8	4.0
	Mathematics	68,198	0	50	25.5	11.9	0.91	3.6	7.2	7.1
	Science & Technology	68,212	0	48	24.0	8.7	0.88	3.0	8.2	6.0
10	English Language Arts	55,613	4	82	47.1	13.3	0.92	3.7	5.1	4.8
	Mathematics	61,297	0	60	23.9	13.3	0.93	3.6	6.9	6.5
	Science & Technology	60,517	0	57	25.4	11.2	0.91	3.3	5.1	5.4

RELIABILITY OF PERFORMANCE LEVEL CATEGORIZATION

All test scores contain measurement error; thus classifications based on test scores are also subject to measurement error. After the performance levels were specified and students were classified into those levels, empirical analyses were conducted to determine the statistical accuracy and consistency of the classifications.

Accuracy

Accuracy refers to the extent to which decisions based on test scores match decisions that would have been made if the scores did not contain any measurement error. Accuracy must be estimated because errorless test scores do not exist.

Consistency

Consistency measures the extent to which classification decisions based on test scores match the decisions based on scores from a second, parallel, form of the same test. Consistency can be evaluated directly from actual responses to test questions if two complete, parallel, forms of the test are given to the same group of students. This is usually impractical, especially on lengthy tests such as the MCAS tests. To overcome this issue, techniques have been developed to estimate both accuracy and consistency of classification decisions based on a single administration of a test. The technique developed by Livingston and Lewis (1995) was used for the MCAS tests because their technique can be used with both constructed-response and multiple-choice questions.

Calculating Accuracy

All of the accuracy and consistency estimation techniques described below make use of the concept of "true scores" in the sense of classical test theory. A true score is the score that would be obtained on a test that had no measurement error. It is a theoretical concept that cannot be observed, although it can be estimated. Following Livingston and Lewis (1995), the true-score distribution for the MCAS tests was estimated using a four-parameter beta distribution, which is a flexible model that allows for extreme degrees of skewness in test scores.

In the Livingston and Lewis method, the estimated "true scores" are used to classify students into their "true" performance category, which is labeled "true status." After various technical adjustments (which are described in Livingston and Lewis, 1995), a 4 × 4 contingency table is created for each test and grade level. The cells in the table are the proportion of students who were classified into each performance category by the actual (or observed) scores on MCAS (i.e., observed status) and by the "true scores" (i.e., "true status"). As an example, Table 14-2 shows the accuracy contingency table for fourth-grade English Language Arts. The accuracy contingency tables for all grades and subjects are provided in Appendix C (under step 5). Additional steps in the analysis are also shown in Appendix C.

Table 14-2
Accuracy Contingency Table for Grade 4 English Language Arts

True Status	Observed Status			
	Failing	Needs Improvement	Proficient	Advanced
Failing	**.11**	.02	.00	.00
Needs Improvement	.04	**.62**	.04	.00
Proficient	.00	.02	**.15**	.00
Advanced	.00	.00	.00	**.00**

Proportions on the diagonal (in bold) indicate exact agreement between the observed status and "true status." If the test were perfectly accurate, all of the off-diagonal cells would be zero. Accuracy is the sum of the diagonal (i.e., the proportion of exact agreement across the four performance levels). In Table 14-2, the diagonal sums to .88, indicating that 88 percent of the students were classified into exactly the same performance categories by their observed scores and their "true scores."

Kappa

Another way to measure consistency is to use Cohen's (1960) coefficient κ (kappa), which assesses the proportion of consistent classifications after removing the proportion of consistent classification that would be expected by chance. Cohen's κ can be used to estimate the classification consistency of a test from two parallel forms of the test. The second form in this case was the one estimated using the Livingston and Lewis (1995) method. Cohen's κ is shown in Table 14-3. Because κ is corrected for chance, the values of κ are lower than the other consistency estimates in Table 14-3.

Calculating Consistency

To estimate consistency, the "true scores" are used to estimate the distribution of classifications on an independent, parallel test form. After statistical adjustments (see Livingston and Lewis, 1995), a new 4 × 4 contingency table is created for each test and grade level that shows the proportion of students who were classified into each performance category by the actual test and by another (hypothetical) parallel test form. Consistency, which is the proportion of students classified into exactly the same categories by the two forms of the test, is the sum of the diagonal for the new contingency table. The consistency contingency tables are shown under step 7 in Appendix C.

Results of Accuracy, Consistency, and Kappa Analyses

The accuracy, consistency, and kappa indices for all grades and subjects are summarized in Table 14-3.

Table 14-3
Estimates of Accuracy and Consistency of Performance Level Classification

Grade	Subject	Accuracy	Consistency	Kappa (κ)
4	English Language Arts	.88	.83	.65
	Mathematics	.77	.68	.54
	Science & Technology	.78	.69	.51
8	English Language Arts	.80	.73	.57
	Mathematics	.79	.71	.58
	Science & Technology	.77	.68	.53
10	English Language Arts	.81	.73	.62
	Mathematics	.82	.75	.61
	Science & Technology	.82	.74	.61

Another way of evaluating accuracy is to estimate the probability of students being classified as being in a particular performance-level category, given that their "true status" was that same category. For example, what is the probability that students who are *really* Proficient (based on their theoretical "true score") will be classified as Proficient based on their MCAS scores? Table 14-4 shows these estimated probabilities.

Table 14-4 Estimated Probability of Being Classified at a Proficiency Level Given that the "True Status" is that Level					
Grade	Subject	Failing	Needs Improvement	Proficient	Advanced
4	English Language Arts	.82	.89	.86	.56
	Mathematics	.83	.77	.70	.80
	Science & Technology	.84	.75	.80	.71
8	English Language Arts	.82	.65	.93	.68
	Mathematics	.90	.67	.74	.80
	Science & Technology	.85	.65	.83	.62
10	English Language Arts	.83	.74	.88	.72
	Mathematics	.92	.68	.71	.81
	Science & Technology	.86	.79	.82	.56

For certain decisions, concern may be highest regarding decisions made about a particular threshold. For example, if a college gave credit to students who achieved an Advanced Placement test score of four or five, but not one, two, or three, one might be interested in the accuracy of the dichotomous decision, below four versus four or above. Table 14-5 reports accuracy and consistency for various dichotomous categorizations on MCAS.

Table 14-5 Accuracy and Consistency of Dichotomous Categorizations							
Grade	Subject	Accuracy			Consistency		
		F/NI	NI/P	P/A	F/NI	NI/P	P/A
4	English Language Arts	.94	.94	.995	.92	.92	.99
	Mathematics	.91	.91	.95	.87	.87	.93
	Science & Technology	.95	.87	.96	.93	.82	.93
8	English Language Arts	.92	.89	.99	.90	.86	.97
	Mathematics	.91	.92	.96	.88	.89	.94
	Science & Technology	.88	.90	.99	.84	.86	.98
10	English Language Arts	.92	.91	.98	.89	.88	.96
	Mathematics	.92	.93	.97	.88	.91	.96
	Science & Technology	.91	.92	.99	.87	.89	.98

BEST COPY AVAILABLE

CHAPTER 15
VALIDITY

As noted in the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1985, p. 9), "validity is the most important consideration in test evaluation." Validity refers to whether specific inferences made from test scores are appropriate, meaningful, and useful. There are several types of validity-related evidence that can be used to support appropriate, meaningful, and useful inferences based on test scores.

CONTENT-RELATED EVIDENCE

As noted in the *Standards* (p. 10), evidence of test validity begins with test development and continues throughout the entire testing process. Chapters 2 through 5 of this manual provide ample evidence regarding the alignment between the content of MCAS and the Massachusetts *Curriculum Frameworks.*

RELATIONSHIP BETWEEN MCAS SCORES AND SCORES ON OTHER TESTS

Gong (1999) and Thacker and Hoffman (1999) correlated MCAS scores with scores on the Stanford Achievement Test (SAT-9) and the Metropolitan Achievement Test (MAT-7). Tables 15-1 and 15-2 present examples of their findings. Correlations between similar measures are in boldface. Note, SAT-9 scores are based only on multiple-choice items.

Table 15-1
Correlations Between MCAS and SAT-9 Scores, District A, Grade 4

		SAT-9					MCAS		
		Reading	Language	Composition	Math	Science	ELA	Math	Science & Tech.
SAT-9	Reading								
	Language	.82							
	Composition	.74	.88						
	Math	.76	.77	.69					
	Science	.76	.72	.65	.70				
MCAS	ELA	**.82**	.76	.68	.70	.66			
	Math	.67	.66	.60	**.69**	.61	.74		
	Sci. & Tech.	.71	.65	.59	.64	**.64**	.75	.75	

Table 15-2
Correlations Between MCAS and MAT-7 Scores, District A, Grade 10

		MAT-7					MCAS		
		Reading	Language	Composition	Math	Science	ELA	Math	Science & Tech.
M	Reading								
	Language	.78							
	Composition	.70	.89						

	Math	.75	.74	.67					
	Science	.75	.66	.59	.70				
MCAS	ELA	**.72**	.68	.61	.67	.61			
MCAS	Math	.66	.66	.59	**.81**	.65	.71		
MCAS	Sci. & Tech.	.72	.64	.59	.72	**.71**	.77	.79	

SUBGROUP DIFFERENCES ON MCAS AND OTHER ACHIEVEMENT TESTS

The *Standards for Educational and Psychological Testing* assert that, when possible, validity studies should address subgroups of interest in addition to the entire test-taking population. Differential performance of gender and ethnic subgroups on large-scale assessments has been well documented in the testing literature. A variety of reasons may explain these results, including different course-taking patterns, socioeconomic issues, and students' opportunities to learn. The important question with respect to potential differential validity is not whether subgroup scores differ, but rather whether some aspect of MCAS increases subgroup differences compared to similar tests.

Male-Female Differences

The two MCAS validity studies (Gong, 1999; Thacker and Hoffman, 1999), showed differences between male and female performance on MCAS, as well as on SAT-9 and MAT-7. The differences between male and female students' MCAS scores tended to be minor in both studies. Differences followed the same patterns for MCAS as for scores on SAT-9 and MAT-7. Male students tended to perform slightly better than female students on the mathematics and science and technology portions of all tests and female students performed slightly better than male students on the reading and writing portions of the tests. Statistical analysis of the results showed no significant differences between the MCAS, SAT-9, and MAT-7 in terms of gender differences.

Ethnic Group Differences

Larger differences in mean MCAS, SAT-9, and MAT-7 scores were found across ethnic subgroups. Both studies (Gong, 1999; Thacker & Hoffman, 1999) indicated that MCAS is similar to the other tests with respect to mean score differences across ethnic subgroups. Thacker and Hoffman (1999) found ethnicity differences small compared to differences due to course-taking patterns. For example, when predicting grade 10 MCAS science and technology scores from MAT-7 science scores, accounting for the courses the students took improved the r-square from .55 to .61. Adding ethnicity to MAT-7 scores and courses taken did not further improve the r-square. Findings in mathematics were similar.

SECTION V
REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1985). *Standards for educational and psychological testing*. Washington, DC: APA.

Brown, F. G. (1983). *Principles of educational and psychological testing* (3rd Edition). Fort Worth: Holt, Rinehart and Winston.

Code of Fair Testing Practices in Education. (1988) Washington, DC: Joint Committee on Testing Practices. (Mailing Address: Joint Committee on Testing Practices, American Psychological Association, 750 First Avenue, NE, Washington, DC 20002-4242.)

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46.

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334.

Dorans, N. J., & Holland, P.W. (1993). DIF detection and description. In P.W. Holland & H. Wainer (Eds.) *Differential item functioning* (pp. 35–66). Hillsdale, N.J.: Lawrence Erlbaum Associates.

Dorans, N.J., & Kulick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement*, 23, 355–368.

Gong, Brian (1999). Relationships between student performance on the MCAS (Massachusetts Comprehensive Assessment System) and other tests—collaborating district A, grades 4 and 10. prepared for the Massachusetts Department of Education. Dover, N.H. The National Center for the Improvement of Educational Assessment, Inc.

Livingston, S.A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement*, 32, 179–197.

Massachusetts Department of Education (1997a), *English Language Arts Curriculum Framework*, Malden, Mass.

Massachusetts Department of Education (1997b), *Mathematics Curriculum Framework: Achieving Mathematical Power*, Malden, Mass.

Massachusetts Department of Education (1997c), *Science and Technology Curriculum Framework: Owning the Questions Through Science and Technology*, Malden, Mass.

Massachusetts Department of Education (1998a). *The Massachusetts Comprehensive Assessment System: Release of May 1998 Test Items*, Malden, Mass.

Massachusetts Department of Education (1998b), *Guide to the Massachusetts Comprehensive Assessment System: English Language Arts*, Malden, Mass.

Massachusetts Department of Education (1998c), *Guide to the Massachusetts Comprehensive Assessment System: Mathematics, Malden*, Mass.

Massachusetts Department of Education (1998d), *Guide to the Massachusetts Comprehensive Assessment System: Science and Technology*, Malden, Mass.

Massachusetts Department of Education (1998e), *Principals Administration Manual*, Malden, Mass.

Massachusetts Department of Education (1998f), *Test Administrators Manual*, Malden, Mass.

Massachusetts Department of Education (1998g), *The Massachusetts Comprehensive Assessment System: Requirements for Test Scheduling, Student Participation, and Test Security and Ethics*, Malden.

BEST COPY AVAILABLE

Massachusetts Department of Education (1998h), *Understanding Your MCAS 1998 Student Report for Parents/Guardians*, Malden, Mass.

Massachusetts Department of Education (1998i), *Guide to Interpreting the 1998 MCAS School and District Reports*, Malden, Mass.

Massachusetts Department of Education, (1998j), *1998 Statewide Summary of District Performance on the Massachusetts Comprehensive Assessment System (MCAS)*, Malden, Mass.

Massachusetts Department of Education, (1998k), *Report of 1998 Statewide Results: The Massachusetts Comprehensive Assessment System (MCAS)*, Malden, Mass.

Thacker, Arthur A. and Hoffman, R. Gene (1999). Relationship between MCAS and SAT-9 for one district in Massachusetts. (Report No. FR-WATSD-99-05). Radcliff, Ky.: HumRRO.

APPENDIX A

Appendix A: MCAS Committee Members

English Language Arts Assessment Development Committee

Bill Amorosi	Belmonte Middle School, Saugus Public Schools
Sorel Berman	Brookline High School, Brookline Public Schools
Ann Connolly-Tolkoff	City on a Hill Charter School (Boston)
Anne Graham	Galvin Middle School, Wakefield Public Schools
Yvonne Gunzburger	Hemenway Elementary, Framingham Public Schools
Susan Horn	Adams-Cheshire Regional Public Schools
William Irvin	Pittsfield Public Schools
Shirley Kountze	Brooks/Hobbs Magnet School, Medford Public Schools
James McDermott	South High Community School, Worcester Public Schools
Laurie Palmer	Memorial School, Natick Public Schools
Lorraine Plasse	Springfield Public Schools
David Roach	Millbury Public Schools
Anne Steele	Shrewsbury High School, Shrewsbury Public Schools
Sandra Stotsky	Harvard Graduate School of Education/Boston University
George Viglirolo	Brookline High School, Brookline Public Schools
Robert Zeeb	Newton Public Schools

Mathematics Assessment Development Committee

Jim Alberque	Worcester State College
Brian Barnes	Mansfield High School, Mansfield Public Schools
Peg Bondorew	Northeastern University
Maureen Chapman-Fahey	Medford Public Schools
David Daniels	Longmeadow High School, Longmeadow Public Schools
William Day	Lawrence School, Falmouth Public Schools
Hal Dickert	Hopkinton Middle School, Hopkinton Public Schools
Paul Donovan	Blue Hills Regional Technical High School
Barbara Haig	Marion Zeh School, Northborough Public Schools
Marcia Harol	Andover High School, Andover Public Schools
Maggi Hartnett	Ayer Senior High School, Ayer Public Schools
Patricia Hills	Holyoke Public Schools
Carol Hynes	Leominster High School, Leominster Public Schools
Joan Kenney	Harvard Graduate School of Education
Deborah King	Monatiquot School, Braintree Public Schools
Michele Kingsland-Smith	Ahern Middle School, Foxborough Public Schools
Raynold Lewis	Worcester Technical Institute
Gloria Moran	M.G. Williams Junior High, Bridgewater-Raynham School District
Donna Pappalardo	Parker Middle School, Reading Public Schools
Christine Redford	Joshua Eaton Elementary, Reading Public Schools
Guy Roy	Plymouth Public Schools
Bernard Ryder	Agawam Public Schools (Retired)
Donna Scanlon	Holyoke Public Schools

Margaret Skowron	Highland Elementary School, Brimfield Public Schools
Nancy Sprague	Bridgewater State College
Kathy VanCamp	Brimfield Elementary School, Brimfield Public Schools
Nancy Zamarro	Worcester Vocational High School, Worcester Public Schools
Giselle Zangari	Boston University Academy

SCIENCE & TECHNOLOGY ASSESSMENT DEVELOPMENT COMMITTEE

Althea Brown	Medford High School, Medford Public Schools
Kathleen Brown	Hudson Public Schools
Paul Cavanagh	North Attleborough High School, North Attleborough Public Schools
Mary Corcoran	Massachusetts Association of Science Supervisors
Charles Corley	McCall Middle School, Winchester Public Schools
Mary Creed	Fall River Public Schools
Joyce Croce	Tyngsborough Public Schools
Howard Dimmick	Stoneham Public Schools
Susan Ferguson-Ellia	Oxford Middle School, Oxford Public Schools
John Fusco	Winchester High School, Winchester Public Schools
Bradford George	Hale Middle School, Stowe Public Schools
Ilia Gonzalez Alonso	Cambridge Public Schools
Diane Goodman	Alfred Zanetti School, Springfield Public Schools
James Hamos	University of Massachusetts Medical Center
Michael Lewandowski	Joseph Case High School, Swansea Public Schools
Michael Lynch	Andover High School, Andover Public Schools
Patrick Markham	Pittsfield Public Schools
Maureen Moir	Bridgewater State College
Louise Mary Nolan	Woburn Public Schools
Maxine Rosenberg	Newton Public Schools
Robert Sartwell	Agawam Public Schools
Peter Shaughnessy	Northampton High School, Northampton Public Schools
Pamela Tickle	FallRiver Public Schools
Maria Torres	Fitchburg Public Schools
Mike Zapantis	Pittsfield Public Schools

STATEWIDE ASSESSMENT ADVISORY COMMITTEE

James Argir	Massachusetts Elementary School Principals Association
Guessippina Bonner	Massachusetts Teachers Association
MaryAnn Byrnes	Consultant on Special Education
Mary Campbell	Horace Mann School for the Deaf, Boston Public Schools
Jim Caradonio	Worcester Public Schools
John Cawthorne	School of Education, Boston College
John Collins	Holy Cross College
Kathleen Conole	Greater Lowell Regional Vocational Technical School
Ruth Ann Corbin	Massachusetts Vocational Association
June Coutu	Massachusetts Council for the Social Studies
Maryellen Donahue	Boston Public Schools
David Fredette	Massachusetts Council of Teachers of English
Al Galante	Massachusetts Association of Teachers of Mathematics

BEST COPY AVAILABLE

Lorraine Greiff	Massachusetts Office on Disability
Ellen Guiney	Boston Plan for Excellence
William Irvin	Pittsfield Public Schools
Julia Landau	Massachusetts Advocacy Center
Yu-Lan Lin	Massachusetts Foreign Language Association
Charles E. Martin, Jr.	Rockport Public Schools
Louise Mary Nolan	Massachusetts Association of School Superintendents
Stephen H. Pronovost	Massachusetts Secondary School Administrators Association
F. Paul Quatromoni	Massachusetts Association of Science Supervisors
Angel G. Ramirez, Jr.	Massachusetts Association of School Committees
Jonathan Rappaport	Worcester Public Schools
Jack Rennie	Massachusetts Business Alliance
Paul Reville	Harvard Graduate School of Education
Roger L. Rice	Multicultural Education Training and Advocacy
Dennis Richards	Reading Public Schools
Connie Rizoli	House Committee on Education, Massachusetts State Legislature
Gregory T. Scotten	Massachusetts Secondary School Administrators Association
Frank Vacirca	Massachusetts Association of Vocational Administrators
Brendan Walsh	Council of Administrators of Compensatory Education

English Language Learner Focus Group

Bethel Bilezikian Charkoudian	English High School, Boston Public Schools
Mary Cazabon	Cambridge Public Schools
Marguerite Goes	Lowell Public Schools
Georgette Gonsalves	Boston Public Schools
Mary Ann Lachat	The Regional Lab, Brown University
Jill McCarthy	Newton Public Schools
Susan J. McGilvray-Rivet	Framingham Public Schools
Marla Perez-Selles	Cambridge Public Schools
Kay Polga	Brookline Public Schools
Rosalie Porter	READ Institute
Roger Rice	Multicultural Education Training and Advocacy
Kathryn L. Riley	Consultant on Bilingual Education
Connie Rizoli	House Committee on Education, Massachusetts State Legislature

Special Education Focus Group

Michael Bello	Learning Center for Deaf Children
MaryAnn Byrnes	Consultant on Special Education
Joan DeGeorge Schirmer	West Bridgewater Public Schools
Cynthia Essex	Perkins School for the Blind
Julia Landau	Massachusetts Advocacy Center
Katherine Levine	East Bridgewater Public Schools
William H. Marginson	New Bedford Schools
Tom Miller	Perkins School for the Blind
Lorna Nickerson Kaufman	Kaufman Associates
Suzanne Recane	Learning Center for Deaf Children

David Riley	Massachusetts Urban Project
Richard Robison	Federation for Children
Tim Sindelar	Disability Law Center
Joanne Testaverdi	Northeast Regional Vocational School

BIAS REVIEW COMMITTEE MEMBERS

Anthony Baxter	Salem State College
Gwenn Blackburn	Medford High School, Medford Public Schools
Guessippina Bonner	Massachusetts Teachers Association
Cathleen Boynton	Brockton High School, Brockton Public Schools
Althea Brown	Medford High School, Medford Public Schools
Kriner Cash	Vineyard Haven Public Schools
Kerry Cavallaro	Norfolk County Regional Vocational High School
John Cawthorne	School of Education, Boston College
Veronica Griffin	Worcester Public Schools
Sumru Erkur	Stone Center, Wellesley College
Carol House	Cambridge Rindge and Latin School, Cambridge Public Schools
Deidre Loughlin	Worcester Public Schools
Paula Martin	Needham Public Schools
Fern Marx	Stone Center, Wellesley College
Margarita Poles	Parent, Boston Public Schools
Lionel Reinford	Worcester Public Schools
Wanda S. Franklin	Needham Public Schools
Meg Wilder Watson	Boston Public Schools

NATIONAL TECHNICAL ADVISORY COMMITTEE

Ron Hambleton	University of Massachusetts, Amherst
George Madaus	Center for the Study of Testing, Evaluation and Educational Policy, Boston College
Barbara Plake	University of Nebraska
Doug Rindone	Connecticut Department of Education
Roger Trent	Ohio Department of Education

APPENDIX B

Grade 4 Reading.xls MC Common

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	0	0.00%	0.00%
0.00 to 0.09	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to 0.19	0	0.00%	0.00%	0	0.00%	0.00%
0.20 to 0.29	0	0.00%	0.00%	3	10.71%	10.71%
0.30 to 0.39	1	3.57%	3.57%	14	50.00%	60.71%
0.40 to 0.49	5	17.86%	21.43%	9	32.14%	92.86%
0.50 to 0.59	8	28.57%	50.00%	2	7.14%	100.00%
0.60 to 0.69	6	21.43%	71.43%	0	0.00%	100.00%
0.70 to 0.79	6	21.43%	92.86%	0	0.00%	100.00%
0.80 to 0.89	2	7.14%	100.00%	0	0.00%	100.00%
0.90 to 0.99	0	0.00%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**28**			**28**	
Average		**0.61**			**0.38**	
Std. Dev.		**0.134806838**			**0.075861891**	
Minimum		**0.31**			**0.20**	
Maximum		**0.88**			**0.50**	
Range		**0.57**			**0.30**	

Grade 4 Reading.xls MC Matrix

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	0	0.00%	0.00%
0.00 to 0.09	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to 0.19	0	0.00%	0.00%	3	3.13%	3.13%
0.20 to 0.29	1	1.04%	1.04%	18	18.75%	21.88%
0.30 to 0.39	10	10.42%	11.46%	35	36.46%	58.33%
0.40 to 0.49	14	14.58%	26.04%	38	39.58%	97.92%
0.50 to 0.59	20	20.83%	46.88%	2	2.08%	100.00%
0.60 to 0.69	17	17.71%	64.58%	0	0.00%	100.00%
0.70 to 0.79	22	22.92%	87.50%	0	0.00%	100.00%
0.80 to 0.89	10	10.42%	97.92%	0	0.00%	100.00%
0.90 to 0.99	2	2.08%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**96**			**96**	
Average		**0.61**			**0.36**	
Std. Dev.		**0.159324093**			**0.084429717**	
Minimum		**0.26**			**0.13**	
Maximum		**0.93**			**0.54**	
Range		**0.67**			**0.41**	

Grade 4 Reading.xls OR

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	0	0.00%	0.00%
0.00 to 0.09	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to 0.19	0	0.00%	0.00%	0	0.00%	0.00%
0.20 to 0.29	2	6.90%	6.90%	0	0.00%	0.00%
0.30 to 0.39	8	27.59%	34.48%	1	3.45%	3.45%
0.40 to 0.49	9	31.03%	65.52%	14	48.28%	51.72%
0.50 to 0.59	8	27.59%	93.10%	13	44.83%	96.55%
0.60 to 0.69	2	6.90%	100.00%	1	3.45%	100.00%
0.70 to 0.79	0	0.00%	100.00%	0	0.00%	100.00%
0.80 to 0.89	0	0.00%	100.00%	0	0.00%	100.00%
0.90 to 0.99	0	0.00%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**29**			**29**	
Average		**0.44**			**0.49**	
Std. Dev.		**0.10796882**			**0.059963043**	
Minimum		**0.21**			**0.32**	
Maximum		**0.63**			**0.61**	
Range		**0.43**			**0.29**	

Grade 4 Reading.xls MC

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	0	0.00%	0.00%
0.00 to 0.09	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to 0.19	0	0.00%	0.00%	3	2.42%	2.42%
0.20 to 0.29	1	0.81%	0.81%	21	16.94%	19.35%
0.30 to 0.39	11	8.87%	9.68%	49	39.52%	58.87%
0.40 to 0.49	19	15.32%	25.00%	47	37.90%	96.77%
0.50 to 0.59	28	22.58%	47.58%	4	3.23%	100.00%
0.60 to 0.69	23	18.55%	66.13%	0	0.00%	100.00%
0.70 to 0.79	28	22.58%	88.71%	0	0.00%	100.00%
0.80 to 0.89	12	9.68%	98.39%	0	0.00%	100.00%
0.90 to 0.99	2	1.61%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**124**			**124**	
Average		**0.61**			**0.36**	
Std. Dev.		**0.153607614**			**0.082662551**	
Minimum		**0.26**			**0.13**	
Maximum		**0.93**			**0.54**	
Range		**0.67**			**0.41**	

Grade 8 Mathematics.xls MC Matrix

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	1	1.67%	1.67%
0.00 to 0.09	0	0.00%	0.00%	1	1.67%	3.33%
0.10 to 0.19	0	0.00%	0.00%	2	3.33%	6.67%
0.20 to 0.29	1	1.67%	1.67%	12	20.00%	26.67%
0.30 to 0.39	11	18.33%	20.00%	20	33.33%	60.00%
0.40 to 0.49	16	26.67%	46.67%	22	36.67%	96.67%
0.50 to 0.59	12	20.00%	66.67%	2	3.33%	100.00%
0.60 to 0.69	8	13.33%	80.00%	0	0.00%	100.00%
0.70 to 0.79	10	16.67%	96.67%	0	0.00%	100.00%
0.80 to 0.89	1	1.67%	98.33%	0	0.00%	100.00%
0.90 to 0.99	1	1.67%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**60**			**60**	
Average		**0.53**			**0.35**	
Std. Dev.		**0.158547343**			**0.10628538**	
Minimum		**0.26**			**-0.05**	
Maximum		**0.93**			**0.57**	
Range		**0.67**			**0.62**	

Grade 8 Mathematics.xls MC Common

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	0	0.00%	0.00%
0.00 to 0.09	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to 0.19	0	0.00%	0.00%	0	0.00%	0.00%
0.20 to 0.29	2	9.52%	9.52%	5	23.81%	23.81%
0.30 to 0.39	1	4.76%	14.29%	7	33.33%	57.14%
0.40 to 0.49	2	9.52%	23.81%	9	42.86%	100.00%
0.50 to 0.59	7	33.33%	57.14%	0	0.00%	100.00%
0.60 to 0.69	3	14.29%	71.43%	0	0.00%	100.00%
0.70 to 0.79	3	14.29%	85.71%	0	0.00%	100.00%
0.80 to 0.89	3	14.29%	100.00%	0	0.00%	100.00%
0.90 to 0.99	0	0.00%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**21**			**21**	
Average		**0.58**			**0.36**	
Std. Dev.		**0.162130165**			**0.076188988**	
Minimum		**0.29**			**0.23**	
Maximum		**0.86**			**0.49**	
Range		**0.57**			**0.26**	

Grade 8 Mathematics.xls MC

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	1	1.23%	1.23%
0.00 to 0.09	0	0.00%	0.00%	1	1.23%	2.47%
0.10 to 0.19	0	0.00%	0.00%	2	2.47%	4.94%
0.20 to 0.29	3	3.70%	3.70%	17	20.99%	25.93%
0.30 to 0.39	12	14.81%	18.52%	27	33.33%	59.26%
0.40 to 0.49	18	22.22%	40.74%	31	38.27%	97.53%
0.50 to 0.59	19	23.46%	64.20%	2	2.47%	100.00%
0.60 to 0.69	11	13.58%	77.78%	0	0.00%	100.00%
0.70 to 0.79	13	16.05%	93.83%	0	0.00%	100.00%
0.80 to 0.89	4	4.94%	98.77%	0	0.00%	100.00%
0.90 to 0.99	1	1.23%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**81**			**81**	
Average		**0.54**			**0.35**	
Std. Dev.		**0.159722077**			**0.099097938**	
Minimum		**0.26**			**-0.05**	
Maximum		**0.93**			**0.57**	
Range		**0.67**			**0.62**	

Grade 8 Mathematics.xls SA

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	0	0.00%	0.00%
0.00 to 0.09	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to 0.19	0	0.00%	0.00%	0	0.00%	0.00%
0.20 to 0.29	2	11.76%	11.76%	0	0.00%	0.00%
0.30 to 0.39	2	11.76%	23.53%	2	11.76%	11.76%
0.40 to 0.49	5	29.41%	52.94%	6	35.29%	47.06%
0.50 to 0.59	3	17.65%	70.59%	9	52.94%	100.00%
0.60 to 0.69	2	11.76%	82.35%	0	0.00%	100.00%
0.70 to 0.79	2	11.76%	94.12%	0	0.00%	100.00%
0.80 to 0.89	1	5.88%	100.00%	0	0.00%	100.00%
0.90 to 0.99	0	0.00%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**17**			**17**	
Average		**0.52**			**0.49**	
Std. Dev.		**0.176176716**			**0.080718283**	
Minimum		**0.23**			**0.31**	
Maximum		**0.86**			**0.59**	
Range		**0.63**			**0.28**	

Grade 8 Mathematics.xls OR

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	0	0.00%	0.00%
0.00 to 0.09	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to 0.19	0	0.00%	0.00%	0	0.00%	0.00%
0.20 to 0.29	4	22.22%	22.22%	0	0.00%	0.00%
0.30 to 0.39	6	33.33%	55.56%	0	0.00%	0.00%
0.40 to 0.49	5	27.78%	83.33%	0	0.00%	0.00%
0.50 to 0.59	3	16.67%	100.00%	3	16.67%	16.67%
0.60 to 0.69	0	0.00%	100.00%	12	66.67%	83.33%
0.70 to 0.79	0	0.00%	100.00%	3	16.67%	100.00%
0.80 to 0.89	0	0.00%	100.00%	0	0.00%	100.00%
0.90 to 0.99	0	0.00%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**18**			**18**	
Average		**0.38**			**0.64**	
Std. Dev.		**0.101535294**			**0.04633284**	
Minimum		**0.21**			**0.55**	
Maximum		**0.55**			**0.71**	
Range		**0.35**			**0.16**	

Grade 4 Science.xls MC Matrix

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	0	0.00%	0.00%
0.00 to 0.09	0	0.00%	0.00%	2	2.78%	2.78%
0.10 to 0.19	1	1.39%	1.39%	7	9.72%	12.50%
0.20 to 0.29	0	0.00%	1.39%	18	25.00%	37.50%
0.30 to 0.39	3	4.17%	5.56%	29	40.28%	77.78%
0.40 to 0.49	9	12.50%	18.06%	16	22.22%	100.00%
0.50 to 0.59	14	19.44%	37.50%	0	0.00%	100.00%
0.60 to 0.69	19	26.39%	63.89%	0	0.00%	100.00%
0.70 to 0.79	13	18.06%	81.94%	0	0.00%	100.00%
0.80 to 0.89	11	15.28%	97.22%	0	0.00%	100.00%
0.90 to 0.99	2	2.78%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**72**			**72**	
Average		**0.64**			**0.32**	
Std. Dev.		**0.156426246**			**0.095768758**	
Minimum		**0.12**			**0.07**	
Maximum		**0.90**			**0.47**	
Range		**0.78**			**0.40**	

Grade 4 Science.xls MC Common

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	0	0.00%	0.00%
0.00 to 0.09	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to 0.19	0	0.00%	0.00%	1	3.85%	3.85%
0.20 to 0.29	0	0.00%	0.00%	7	26.92%	30.77%
0.30 to 0.39	2	7.69%	7.69%	10	38.46%	69.23%
0.40 to 0.49	2	7.69%	15.38%	8	30.77%	100.00%
0.50 to 0.59	4	15.38%	30.77%	0	0.00%	100.00%
0.60 to 0.69	10	38.46%	69.23%	0	0.00%	100.00%
0.70 to 0.79	3	11.54%	80.77%	0	0.00%	100.00%
0.80 to 0.89	4	15.38%	96.15%	0	0.00%	100.00%
0.90 to 0.99	1	3.85%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**26**			**26**	
Average		**0.65**			**0.32**	
Std. Dev.		**0.148294924**			**0.082174299**	
Minimum		**0.35**			**0.10**	
Maximum		**0.93**			**0.46**	
Range		**0.58**			**0.36**	

Grade 4 Science.xls MC

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	0	0.00%	0.00%
0.00 to 0.09	0	0.00%	0.00%	2	2.04%	2.04%
0.10 to 0.19	1	1.02%	1.02%	8	8.16%	10.20%
0.20 to 0.29	0	0.00%	1.02%	25	25.51%	35.71%
0.30 to 0.39	5	5.10%	6.12%	39	39.80%	75.51%
0.40 to 0.49	11	11.22%	17.35%	24	24.49%	100.00%
0.50 to 0.59	18	18.37%	35.71%	0	0.00%	100.00%
0.60 to 0.69	29	29.59%	65.31%	0	0.00%	100.00%
0.70 to 0.79	16	16.33%	81.63%	0	0.00%	100.00%
0.80 to 0.89	15	15.31%	96.94%	0	0.00%	100.00%
0.90 to 0.99	3	3.06%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**98**			**98**	
Average		**0.64**			**0.32**	
Std. Dev.		**0.153620861**			**0.092019731**	
Minimum		**0.12**			**0.07**	
Maximum		**0.93**			**0.47**	
Range		**0.81**			**0.40**	

Grade 4 Science.xls OR

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	0	0.00%	0.00%
0.00 to 0.09	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to 0.19	0	0.00%	0.00%	0	0.00%	0.00%
0.20 to 0.29	2	11.11%	11.11%	1	5.56%	5.56%
0.30 to 0.39	2	11.11%	22.22%	5	27.78%	33.33%
0.40 to 0.49	7	38.89%	61.11%	8	44.44%	77.78%
0.50 to 0.59	6	33.33%	94.44%	4	22.22%	100.00%
0.60 to 0.69	0	0.00%	94.44%	0	0.00%	100.00%
0.70 to 0.79	1	5.56%	100.00%	0	0.00%	100.00%
0.80 to 0.89	0	0.00%	100.00%	0	0.00%	100.00%
0.90 to 0.99	0	0.00%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**18**			**18**	
Average		**0.46**			**0.43**	
Std. Dev.		**0.122554926**			**0.081041021**	
Minimum		**0.22**			**0.22**	
Maximum		**0.79**			**0.53**	
Range		**0.57**			**0.31**	

Grade 8 Reading.xls MC Matrix

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	0	0.00%	0.00%
0.00 to 0.09	0	0.00%	0.00%	1	1.04%	1.04%
0.10 to 0.19	0	0.00%	0.00%	2	2.08%	3.13%
0.20 to 0.29	0	0.00%	0.00%	6	6.25%	9.38%
0.30 to 0.39	2	2.08%	2.08%	47	48.96%	58.33%
0.40 to 0.49	9	9.38%	11.46%	36	37.50%	95.83%
0.50 to 0.59	19	19.79%	31.25%	4	4.17%	100.00%
0.60 to 0.69	21	21.88%	53.13%	0	0.00%	100.00%
0.70 to 0.79	33	34.38%	87.50%	0	0.00%	100.00%
0.80 to 0.89	12	12.50%	100.00%	0	0.00%	100.00%
0.90 to 0.99	0	0.00%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**96**			**96**	
Average		**0.66**			**0.37**	
Std. Dev.		**0.124689747**			**0.07614902**	
Minimum		**0.37**			**0.07**	
Maximum		**0.89**			**0.53**	
Range		**0.52**			**0.46**	

Grade 8 Reading.xls MC Common

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	0	0.00%	0.00%
0.00 to 0.09	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to 0.19	0	0.00%	0.00%	2	7.14%	7.14%
0.20 to 0.29	0	0.00%	0.00%	7	25.00%	32.14%
0.30 to 0.39	0	0.00%	0.00%	11	39.29%	71.43%
0.40 to 0.49	2	7.14%	7.14%	8	28.57%	100.00%
0.50 to 0.59	1	3.57%	10.71%	0	0.00%	100.00%
0.60 to 0.69	12	42.86%	53.57%	0	0.00%	100.00%
0.70 to 0.79	11	39.29%	92.86%	0	0.00%	100.00%
0.80 to 0.89	2	7.14%	100.00%	0	0.00%	100.00%
0.90 to 0.99	0	0.00%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**28**			**28**	
Average		**0.68**			**0.34**	
Std. Dev.		**0.092892952**			**0.092234891**	
Minimum		**0.42**			**0.15**	
Maximum		**0.81**			**0.49**	
Range		**0.39**			**0.34**	

Grade 8 Reading.xls MC

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	0	0.00%	0.00%
0.00 to 0.09	0	0.00%	0.00%	1	0.81%	0.81%
0.10 to 0.19	0	0.00%	0.00%	4	3.23%	4.03%
0.20 to 0.29	0	0.00%	0.00%	13	10.48%	14.52%
0.30 to 0.39	2	1.61%	1.61%	58	46.77%	61.29%
0.40 to 0.49	11	8.87%	10.48%	44	35.48%	96.77%
0.50 to 0.59	20	16.13%	26.61%	4	3.23%	100.00%
0.60 to 0.69	33	26.61%	53.23%	0	0.00%	100.00%
0.70 to 0.79	44	35.48%	88.71%	0	0.00%	100.00%
0.80 to 0.89	14	11.29%	100.00%	0	0.00%	100.00%
0.90 to 0.99	0	0.00%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**124**			**124**	
Average		**0.66**			**0.37**	
Std. Dev.		**0.118213908**			**0.080894066**	
Minimum		**0.37**			**0.07**	
Maximum		**0.89**			**0.53**	
Range		**0.52**			**0.46**	

Grade 8 Reading.xls OR

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	0	0.00%	0.00%
0.00 to 0.09	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to 0.19	0	0.00%	0.00%	0	0.00%	0.00%
0.20 to 0.29	0	0.00%	0.00%	0	0.00%	0.00%
0.30 to 0.39	0	0.00%	0.00%	0	0.00%	0.00%
0.40 to 0.49	20	68.97%	68.97%	4	13.79%	13.79%
0.50 to 0.59	9	31.03%	100.00%	20	68.97%	82.76%
0.60 to 0.69	0	0.00%	100.00%	5	17.24%	100.00%
0.70 to 0.79	0	0.00%	100.00%	0	0.00%	100.00%
0.80 to 0.89	0	0.00%	100.00%	0	0.00%	100.00%
0.90 to 0.99	0	0.00%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**29**			**29**	
Average		**0.47**			**0.56**	
Std. Dev.		**0.046081857**			**0.042458325**	
Minimum		**0.40**			**0.46**	
Maximum		**0.56**			**0.62**	
Range		**0.16**			**0.16**	

Grade 8 Mathematics.xls MC Matrix

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	1	1.67%	1.67%
0.00 to 0.09	0	0.00%	0.00%	1	1.67%	3.33%
0.10 to 0.19	0	0.00%	0.00%	2	3.33%	6.67%
0.20 to 0.29	1	1.67%	1.67%	12	20.00%	26.67%
0.30 to 0.39	11	18.33%	20.00%	20	33.33%	60.00%
0.40 to 0.49	16	26.67%	46.67%	22	36.67%	96.67%
0.50 to 0.59	12	20.00%	66.67%	2	3.33%	100.00%
0.60 to 0.69	8	13.33%	80.00%	0	0.00%	100.00%
0.70 to 0.79	10	16.67%	96.67%	0	0.00%	100.00%
0.80 to 0.89	1	1.67%	98.33%	0	0.00%	100.00%
0.90 to 0.99	1	1.67%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**60**			**60**	
Average		**0.53**			**0.35**	
Std. Dev.		**0.158547343**			**0.10628538**	
Minimum		**0.26**			**-0.05**	
Maximum		**0.93**			**0.57**	
Range		**0.67**			**0.62**	

Grade 8 Mathematics.xls MC Common

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	0	0.00%	0.00%
0.00 to 0.09	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to 0.19	0	0.00%	0.00%	0	0.00%	0.00%
0.20 to 0.29	2	9.52%	9.52%	5	23.81%	23.81%
0.30 to 0.39	1	4.76%	14.29%	7	33.33%	57.14%
0.40 to 0.49	2	9.52%	23.81%	9	42.86%	100.00%
0.50 to 0.59	7	33.33%	57.14%	0	0.00%	100.00%
0.60 to 0.69	3	14.29%	71.43%	0	0.00%	100.00%
0.70 to 0.79	3	14.29%	85.71%	0	0.00%	100.00%
0.80 to 0.89	3	14.29%	100.00%	0	0.00%	100.00%
0.90 to 0.99	0	0.00%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**21**			**21**	
Average		**0.58**			**0.36**	
Std. Dev.		**0.162130165**			**0.076188988**	
Minimum		**0.29**			**0.23**	
Maximum		**0.86**			**0.49**	
Range		**0.57**			**0.26**	

Grade 8 Mathematics.xls MC

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	1	1.23%	1.23%
0.00 to 0.09	0	0.00%	0.00%	1	1.23%	2.47%
0.10 to 0.19	0	0.00%	0.00%	2	2.47%	4.94%
0.20 to 0.29	3	3.70%	3.70%	17	20.99%	25.93%
0.30 to 0.39	12	14.81%	18.52%	27	33.33%	59.26%
0.40 to 0.49	18	22.22%	40.74%	31	38.27%	97.53%
0.50 to 0.59	19	23.46%	64.20%	2	2.47%	100.00%
0.60 to 0.69	11	13.58%	77.78%	0	0.00%	100.00%
0.70 to 0.79	13	16.05%	93.83%	0	0.00%	100.00%
0.80 to 0.89	4	4.94%	98.77%	0	0.00%	100.00%
0.90 to 0.99	1	1.23%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**81**			**81**	
Average		**0.54**			**0.35**	
Std. Dev.		**0.159722077**			**0.099097938**	
Minimum		**0.26**			**-0.05**	
Maximum		**0.93**			**0.57**	
Range		**0.67**			**0.62**	

Grade 8 Mathematics.xls SA

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	0	0.00%	0.00%
0.00 to 0.09	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to 0.19	0	0.00%	0.00%	0	0.00%	0.00%
0.20 to 0.29	2	11.76%	11.76%	0	0.00%	0.00%
0.30 to 0.39	2	11.76%	23.53%	2	11.76%	11.76%
0.40 to 0.49	5	29.41%	52.94%	6	35.29%	47.06%
0.50 to 0.59	3	17.65%	70.59%	9	52.94%	100.00%
0.60 to 0.69	2	11.76%	82.35%	0	0.00%	100.00%
0.70 to 0.79	2	11.76%	94.12%	0	0.00%	100.00%
0.80 to 0.89	1	5.88%	100.00%	0	0.00%	100.00%
0.90 to 0.99	0	0.00%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**17**			**17**	
Average		**0.52**			**0.49**	
Std. Dev.		**0.176176716**			**0.080718283**	
Minimum		**0.23**			**0.31**	
Maximum		**0.86**			**0.59**	
Range		**0.63**			**0.28**	

Grade 8 Mathematics.xls OR

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	0	0.00%	0.00%
0.00 to 0.09	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to 0.19	0	0.00%	0.00%	0	0.00%	0.00%
0.20 to 0.29	4	22.22%	22.22%	0	0.00%	0.00%
0.30 to 0.39	6	33.33%	55.56%	0	0.00%	0.00%
0.40 to 0.49	5	27.78%	83.33%	0	0.00%	0.00%
0.50 to 0.59	3	16.67%	100.00%	3	16.67%	16.67%
0.60 to 0.69	0	0.00%	100.00%	12	66.67%	83.33%
0.70 to 0.79	0	0.00%	100.00%	3	16.67%	100.00%
0.80 to 0.89	0	0.00%	100.00%	0	0.00%	100.00%
0.90 to 0.99	0	0.00%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**18**			**18**	
Average		**0.38**			**0.64**	
Std. Dev.		**0.101535294**			**0.04633284**	
Minimum		**0.21**			**0.55**	
Maximum		**0.55**			**0.71**	
Range		**0.35**			**0.16**	

Grade 8 Science.xls MC Common

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	0	0.00%	0.00%
0.00 to 0.09	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to 0.19	0	0.00%	0.00%	1	3.85%	3.85%
0.20 to 0.29	1	3.85%	3.85%	13	50.00%	53.85%
0.30 to 0.39	1	3.85%	7.69%	11	42.31%	96.15%
0.40 to 0.49	6	23.08%	30.77%	1	3.85%	100.00%
0.50 to 0.59	7	26.92%	57.69%	0	0.00%	100.00%
0.60 to 0.69	4	15.38%	73.08%	0	0.00%	100.00%
0.70 to 0.79	5	19.23%	92.31%	0	0.00%	100.00%
0.80 to 0.89	2	7.69%	100.00%	0	0.00%	100.00%
0.90 to 0.99	0	0.00%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**26**			**26**	
Average		**0.57**			**0.29**	
Std. Dev.		**0.15248102**			**0.057179878**	
Minimum		**0.24**			**0.16**	
Maximum		**0.86**			**0.40**	
Range		**0.62**			**0.24**	

Grade 8 Science.xls MC Matrix

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	0	0.00%	0.00%
0.00 to 0.09	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to 0.19	0	0.00%	0.00%	2	2.78%	2.78%
0.20 to 0.29	0	0.00%	0.00%	22	30.56%	33.33%
0.30 to 0.39	5	6.94%	6.94%	34	47.22%	80.56%
0.40 to 0.49	13	18.06%	25.00%	14	19.44%	100.00%
0.50 to 0.59	15	20.83%	45.83%	0	0.00%	100.00%
0.60 to 0.69	14	19.44%	65.28%	0	0.00%	100.00%
0.70 to 0.79	17	23.61%	88.89%	0	0.00%	100.00%
0.80 to 0.89	8	11.11%	100.00%	0	0.00%	100.00%
0.90 to 0.99	0	0.00%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**72**			**72**	
Average		**0.62**			**0.33**	
Std. Dev.		**0.145387548**			**0.069443912**	
Minimum		**0.32**			**0.18**	
Maximum		**0.88**			**0.46**	
Range		**0.56**			**0.28**	

Grade 8 Science.xls MC

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	0	0.00%	0.00%
0.00 to 0.09	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to 0.19	0	0.00%	0.00%	3	3.06%	3.06%
0.20 to 0.29	1	1.02%	1.02%	35	35.71%	38.78%
0.30 to 0.39	6	6.12%	7.14%	45	45.92%	84.69%
0.40 to 0.49	19	19.39%	26.53%	15	15.31%	100.00%
0.50 to 0.59	22	22.45%	48.98%	0	0.00%	100.00%
0.60 to 0.69	18	18.37%	67.35%	0	0.00%	100.00%
0.70 to 0.79	22	22.45%	89.80%	0	0.00%	100.00%
0.80 to 0.89	10	10.20%	100.00%	0	0.00%	100.00%
0.90 to 0.99	0	0.00%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**98**			**98**	
Average		**0.60**			**0.32**	
Std. Dev.		**0.147766349**			**0.068679346**	
Minimum		**0.24**			**0.16**	
Maximum		**0.88**			**0.46**	
Range		**0.64**			**0.30**	

Grade 8 Science.xls OR

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	0	0.00%	0.00%
0.00 to 0.09	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to 0.19	0	0.00%	0.00%	0	0.00%	0.00%
0.20 to 0.29	4	21.05%	21.05%	0	0.00%	0.00%
0.30 to 0.39	9	47.37%	68.42%	0	0.00%	0.00%
0.40 to 0.49	5	26.32%	94.74%	5	26.32%	26.32%
0.50 to 0.59	1	5.26%	100.00%	11	57.89%	84.21%
0.60 to 0.69	0	0.00%	100.00%	2	10.53%	94.74%
0.70 to 0.79	0	0.00%	100.00%	0	0.00%	94.74%
0.80 to 0.89	0	0.00%	100.00%	0	0.00%	94.74%
0.90 to 0.99	0	0.00%	100.00%	0	0.00%	94.74%
1.00	0	0.00%	100.00%	0	0.00%	94.74%
Total Count		**19**			**18**	
Average		**0.37**			**0.54**	
Std. Dev.		**0.079877897**			**0.045489351**	
Minimum		**0.26**			**0.47**	
Maximum		**0.54**			**0.61**	
Range		**0.28**			**0.14**	

Grade 10 Reading.xls MC Matrix

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	0	0.00%	0.00%
0.00 to 0.09	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to 0.19	0	0.00%	0.00%	3	3.13%	3.13%
0.20 to 0.29	0	0.00%	0.00%	19	19.79%	22.92%
0.30 to 0.39	4	4.17%	4.17%	45	46.88%	69.79%
0.40 to 0.49	8	8.33%	12.50%	27	28.13%	97.92%
0.50 to 0.59	27	28.13%	40.63%	2	2.08%	100.00%
0.60 to 0.69	25	26.04%	66.67%	0	0.00%	100.00%
0.70 to 0.79	20	20.83%	87.50%	0	0.00%	100.00%
0.80 to 0.89	10	10.42%	97.92%	0	0.00%	100.00%
0.90 to 0.99	2	2.08%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**96**			**96**	
Average		**0.63**			**0.35**	
Std. Dev.		**0.133244927**			**0.082516346**	
Minimum		**0.32**			**0.16**	
Maximum		**0.94**			**0.53**	
Range		**0.62**			**0.37**	

Grade 10 Reading.xls MC Common

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	0	0.00%	0.00%
0.00 to 0.09	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to 0.19	0	0.00%	0.00%	0	0.00%	0.00%
0.20 to 0.29	0	0.00%	0.00%	7	21.88%	21.88%
0.30 to 0.39	1	3.13%	3.13%	17	53.13%	75.00%
0.40 to 0.49	3	9.38%	12.50%	8	25.00%	100.00%
0.50 to 0.59	6	18.75%	31.25%	0	0.00%	100.00%
0.60 to 0.69	7	21.88%	53.13%	0	0.00%	100.00%
0.70 to 0.79	10	31.25%	84.38%	0	0.00%	100.00%
0.80 to 0.89	4	12.50%	96.88%	0	0.00%	100.00%
0.90 to 0.99	1	3.13%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**32**			**32**	
Average		**0.66**			**0.34**	
Std. Dev.		**0.1329792**			**0.064731922**	
Minimum		**0.39**			**0.22**	
Maximum		**0.91**			**0.44**	
Range		**0.52**			**0.22**	

Grade 10 Reading.xls MC

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	0	0.00%	0.00%
0.00 to 0.09	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to 0.19	0	0.00%	0.00%	3	2.34%	2.34%
0.20 to 0.29	0	0.00%	0.00%	26	20.31%	22.66%
0.30 to 0.39	5	3.91%	3.91%	62	48.44%	71.09%
0.40 to 0.49	11	8.59%	12.50%	35	27.34%	98.44%
0.50 to 0.59	33	25.78%	38.28%	2	1.56%	100.00%
0.60 to 0.69	32	25.00%	63.28%	0	0.00%	100.00%
0.70 to 0.79	30	23.44%	86.72%	0	0.00%	100.00%
0.80 to 0.89	14	10.94%	97.66%	0	0.00%	100.00%
0.90 to 0.99	3	2.34%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**128**			**128**	
Average		**0.64**			**0.35**	
Std. Dev.		**0.133167702**			**0.078291574**	
Minimum		**0.32**			**0.16**	
Maximum		**0.94**			**0.53**	
Range		**0.62**			**0.37**	

Grade 10 Reading.xls OR

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	0	0.00%	0.00%
0.00 to 0.09	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to 0.19	0	0.00%	0.00%	0	0.00%	0.00%
0.20 to 0.29	0	0.00%	0.00%	0	0.00%	0.00%
0.30 to 0.39	6	18.75%	18.75%	0	0.00%	0.00%
0.40 to 0.49	25	78.13%	96.88%	1	3.13%	3.13%
0.50 to 0.59	1	3.13%	100.00%	15	46.88%	50.00%
0.60 to 0.69	0	0.00%	100.00%	16	50.00%	100.00%
0.70 to 0.79	0	0.00%	100.00%	0	0.00%	100.00%
0.80 to 0.89	0	0.00%	100.00%	0	0.00%	100.00%
0.90 to 0.99	0	0.00%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**32**			**32**	
Average		**0.43**			**0.59**	
Std. Dev.		**0.03346947**			**0.038052383**	
Minimum		**0.36**			**0.48**	
Maximum		**0.49**			**0.64**	
Range		**0.13**			**0.16**	

Grade 10 Mathematics.xls MC Common

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	1	3.70%	3.70%
0.00 to 0.09	0	0.00%	0.00%	0	0.00%	3.70%
0.10 to 0.19	0	0.00%	0.00%	1	3.70%	7.41%
0.20 to 0.29	2	7.41%	7.41%	3	11.11%	18.52%
0.30 to 0.39	1	3.70%	11.11%	9	33.33%	51.85%
0.40 to 0.49	6	22.22%	33.33%	10	37.04%	88.89%
0.50 to 0.59	6	22.22%	55.56%	3	11.11%	100.00%
0.60 to 0.69	8	29.63%	85.19%	0	0.00%	100.00%
0.70 to 0.79	3	11.11%	96.30%	0	0.00%	100.00%
0.80 to 0.89	1	3.70%	100.00%	0	0.00%	100.00%
0.90 to 0.99	0	0.00%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**27**			**27**	
Average		**0.55**			**0.37**	
Std. Dev.		**0.152150772**			**0.11968739**	
Minimum		**0.21**			**-0.02**	
Maximum		**0.84**			**0.55**	
Range		**0.63**			**0.57**	

Grade 10 Mathematics.xls MC Matrix

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	0	0.00%	0.00%
0.00 to 0.09	0	0.00%	0.00%	2	2.38%	2.38%
0.10 to 0.19	2	2.38%	2.38%	12	14.29%	16.67%
0.20 to 0.29	12	14.29%	16.67%	31	36.90%	53.57%
0.30 to 0.39	24	28.57%	45.24%	18	21.43%	75.00%
0.40 to 0.49	26	30.95%	76.19%	20	23.81%	98.81%
0.50 to 0.59	11	13.10%	89.29%	1	1.19%	100.00%
0.60 to 0.69	7	8.33%	97.62%	0	0.00%	100.00%
0.70 to 0.79	2	2.38%	100.00%	0	0.00%	100.00%
0.80 to 0.89	0	0.00%	100.00%	0	0.00%	100.00%
0.90 to 0.99	0	0.00%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**84**			**84**	
Average		**0.42**			**0.30**	
Std. Dev.		**0.134536027**			**0.104256169**	
Minimum		**0.13**			**0.05**	
Maximum		**0.79**			**0.50**	
Range		**0.66**			**0.45**	

Grade 10 Mathematics.xls OR

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	0	0.00%	0.00%
0.00 to 0.09	1	3.13%	3.13%	0	0.00%	0.00%
0.10 to 0.19	13	40.63%	43.75%	0	0.00%	0.00%
0.20 to 0.29	8	25.00%	68.75%	0	0.00%	0.00%
0.30 to 0.39	6	18.75%	87.50%	1	3.13%	3.13%
0.40 to 0.49	4	12.50%	100.00%	2	6.25%	9.38%
0.50 to 0.59	0	0.00%	100.00%	5	15.63%	25.00%
0.60 to 0.69	0	0.00%	100.00%	18	56.25%	81.25%
0.70 to 0.79	0	0.00%	100.00%	6	18.75%	100.00%
0.80 to 0.89	0	0.00%	100.00%	0	0.00%	100.00%
0.90 to 0.99	0	0.00%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**32**			**32**	
Average		**0.24**			**0.62**	
Std. Dev.		**0.098530406**			**0.083913838**	
Minimum		**0.09**			**0.38**	
Maximum		**0.44**			**0.74**	
Range		**0.35**			**0.36**	

Grade 10 Mathematics.xls SA

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	0	0.00%	0.00%
0.00 to 0.09	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to 0.19	2	11.76%	11.76%	0	0.00%	0.00%
0.20 to 0.29	2	11.76%	23.53%	1	5.88%	5.88%
0.30 to 0.39	5	29.41%	52.94%	1	5.88%	11.76%
0.40 to 0.49	4	23.53%	76.47%	9	52.94%	64.71%
0.50 to 0.59	1	5.88%	82.35%	5	29.41%	94.12%
0.60 to 0.69	1	5.88%	88.24%	1	5.88%	100.00%
0.70 to 0.79	2	11.76%	100.00%	0	0.00%	100.00%
0.80 to 0.89	0	0.00%	100.00%	0	0.00%	100.00%
0.90 to 0.99	0	0.00%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**17**			**17**	
Average		**0.41**			**0.46**	
Std. Dev.		**0.172336879**			**0.096364046**	
Minimum		**0.18**			**0.23**	
Maximum		**0.75**			**0.62**	
Range		**0.57**			**0.39**	

Grade 10 Mathematics.xls MC

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	1	0.90%	0.90%
0.00 to 0.09	0	0.00%	0.00%	2	1.80%	2.70%
0.10 to 0.19	2	1.80%	1.80%	13	11.71%	14.41%
0.20 to 0.29	14	12.61%	14.41%	34	30.63%	45.05%
0.30 to 0.39	25	22.52%	36.94%	27	24.32%	69.37%
0.40 to 0.49	32	28.83%	65.77%	30	27.03%	96.40%
0.50 to 0.59	17	15.32%	81.08%	4	3.60%	100.00%
0.60 to 0.69	15	13.51%	94.59%	0	0.00%	100.00%
0.70 to 0.79	5	4.50%	99.10%	0	0.00%	100.00%
0.80 to 0.89	1	0.90%	100.00%	0	0.00%	100.00%
0.90 to 0.99	0	0.00%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**111**			**111**	
Average		**0.45**			**0.32**	
Std. Dev.		**0.14959618**			**0.11111187**	
Minimum		**0.13**			**-0.02**	
Maximum		**0.84**			**0.55**	
Range		**0.71**			**0.57**	

Grade 10 Science.xls MC Common

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	0	0.00%	0.00%
0.00 to 0.09	0	0.00%	0.00%	1	3.13%	3.13%
0.10 to 0.19	0	0.00%	0.00%	3	9.38%	12.50%
0.20 to 0.29	2	6.25%	6.25%	9	28.13%	40.63%
0.30 to 0.39	2	6.25%	12.50%	16	50.00%	90.63%
0.40 to 0.49	9	28.13%	40.63%	3	9.38%	100.00%
0.50 to 0.59	2	6.25%	46.88%	0	0.00%	100.00%
0.60 to 0.69	8	25.00%	71.88%	0	0.00%	100.00%
0.70 to 0.79	7	21.88%	93.75%	0	0.00%	100.00%
0.80 to 0.89	2	6.25%	100.00%	0	0.00%	100.00%
0.90 to 0.99	0	0.00%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**32**			**32**	
Average		**0.58**			**0.29**	
Std. Dev.		**0.166461937**			**0.100714684**	
Minimum		**0.21**			**0.03**	
Maximum		**0.84**			**0.48**	
Range		**0.63**			**0.45**	

Grade 10 Science.xls MC Matrix

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	0	0.00%	0.00%
0.00 to 0.09	0	0.00%	0.00%	1	1.04%	1.04%
0.10 to 0.19	0	0.00%	0.00%	12	12.50%	13.54%
0.20 to 0.29	4	4.17%	4.17%	31	32.29%	45.83%
0.30 to 0.39	10	10.42%	14.58%	36	37.50%	83.33%
0.40 to 0.49	22	22.92%	37.50%	16	16.67%	100.00%
0.50 to 0.59	21	21.88%	59.38%	0	0.00%	100.00%
0.60 to 0.69	20	20.83%	80.21%	0	0.00%	100.00%
0.70 to 0.79	16	16.67%	96.88%	0	0.00%	100.00%
0.80 to 0.89	2	2.08%	98.96%	0	0.00%	100.00%
0.90 to 0.99	1	1.04%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**96**			**96**	
Average		**0.55**			**0.30**	
Std. Dev.		**0.148904737**			**0.088332444**	
Minimum		**0.20**			**0.08**	
Maximum		**0.91**			**0.47**	
Range		**0.71**			**0.39**	

Grade 10 Science.xls MC

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	0	0.00%	0.00%
0.00 to 0.09	0	0.00%	0.00%	2	1.56%	1.56%
0.10 to 0.19	0	0.00%	0.00%	15	11.72%	13.28%
0.20 to 0.29	6	4.69%	4.69%	40	31.25%	44.53%
0.30 to 0.39	12	9.38%	14.06%	52	40.63%	85.16%
0.40 to 0.49	31	24.22%	38.28%	19	14.84%	100.00%
0.50 to 0.59	23	17.97%	56.25%	0	0.00%	100.00%
0.60 to 0.69	28	21.88%	78.13%	0	0.00%	100.00%
0.70 to 0.79	23	17.97%	96.09%	0	0.00%	100.00%
0.80 to 0.89	4	3.13%	99.22%	0	0.00%	100.00%
0.90 to 0.99	1	0.78%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**128**			**128**	
Average		**0.56**			**0.30**	
Std. Dev.		**0.153147988**			**0.091272581**	
Minimum		**0.20**			**0.03**	
Maximum		**0.91**			**0.48**	
Range		**0.71**			**0.45**	

Grade 10 Science.xls OR

Range	Difficulty			Pearson Correlation		
	Count	Percentage	Cummulative Percentage	Count	Percentage	Cummulative Percentage
< -0.30	0	0.00%	0.00%	0	0.00%	0.00%
-0.30 to -0.21	0	0.00%	0.00%	0	0.00%	0.00%
-0.20 to -0.11	0	0.00%	0.00%	0	0.00%	0.00%
0.10 to -0.01	0	0.00%	0.00%	0	0.00%	0.00%
0.00 to 0.09	3	9.38%	9.38%	0	0.00%	0.00%
0.10 to 0.19	10	31.25%	40.63%	0	0.00%	0.00%
0.20 to 0.29	9	28.13%	68.75%	0	0.00%	0.00%
0.30 to 0.39	10	31.25%	100.00%	1	3.13%	3.13%
0.40 to 0.49	0	0.00%	100.00%	9	28.13%	31.25%
0.50 to 0.59	0	0.00%	100.00%	20	62.50%	93.75%
0.60 to 0.69	0	0.00%	100.00%	2	6.25%	100.00%
0.70 to 0.79	0	0.00%	100.00%	0	0.00%	100.00%
0.80 to 0.89	0	0.00%	100.00%	0	0.00%	100.00%
0.90 to 0.99	0	0.00%	100.00%	0	0.00%	100.00%
1.00	0	0.00%	100.00%	0	0.00%	100.00%
Total Count		**32**			**32**	
Average		**0.22**			**0.52**	
Std. Dev.		**0.097742538**			**0.067339265**	
Minimum		**0.03**			**0.33**	
Maximum		**0.37**			**0.65**	
Range		**0.34**			**0.32**	

APPENDIX C

Grade 4 English Language Arts.xls
Accuracy and Consistency of Classification

Step 4							
	Predicted Classification (X1)						
True Status	Fail	Needs	Prof	Adv	Marginal		
Fail	**0.10191905**	**0.02288016**	**2.267E-14**	**5.1455E-31**	0.12479922		
Needs	**0.03569928**	**0.62955722**	**0.0316454**	**6.4079E-08**	0.69690197		
Prof	**1.5264E-14**	**0.02241005**	**0.12802861**	**0.00618971**	0.15662836		
Adv	**6.8008E-39**	**4.3873E-10**	**0.00299076**	**0.0186797**	0.02167045		
Marginal	0.13761834	0.67484743	0.16266476	0.02486947	1		
Step 5							
	Actual Classification (X0)						
True Status	Fail	Needs	Prof	Adv	Marginal		
Fail	0.10590606	0.02250555	2.6108E-14	1.213E-31	0.12841161		
Needs	0.03709582	0.61924966	0.03644524	1.5106E-08	0.69279073		
Prof	1.5861E-14	0.02204313	0.14744744	0.00145916	0.17094973		
Adv	7.0669E-39	4.3154E-10	0.00344438	0.00440355	0.00784793		
Marginal	0.14300188	0.66379834	0.18733705	0.00586273	1	accuracy	**0.8770067**
cut1	**0.94039863**		cut2	**0.94151161**		cut3	**0.99509644**
0.10590606	0.02250555		0.78475709	0.03644525		0.99069289	0.00145918
0.03709582	0.83449257		0.02204313	0.15675453		0.00344438	0.00440355
	1			1			1
Step 6							
	X1						
X2	Fail	Needs	Prof	Adv	Marginal		
Fail	0.09629486	0.04132346	1.8925E-08	1.1931E-17	0.13761834		
Needs	0.04132346	0.5953604	0.03815754	6.0289E-06	0.67484743		
Prof	1.8925E-08	0.03815754	0.11793189	0.00657531	0.16266476		
Adv	1.1931E-17	6.0289E-06	0.00657531	0.01828813	0.02486947		
Marginal	0.13761834	0.67484743	0.16266476	0.02486947	1		
Step 7							
	X0						
X2	Fail	Needs	Prof	Adv	Marginal		
Fail	0.10006185	0.04064689	2.1796E-08	2.8127E-18	0.14070876		
Needs	0.04294001	0.58561273	0.04394512	1.4212E-06	0.67249927		
Prof	1.9665E-08	0.0375328	0.13581929	0.00155006	0.17490217		
Adv	1.2398E-17	5.9302E-06	0.00757262	0.00431124	0.0118898		
Marginal	0.14300188	0.66379834	0.18733705	0.00586273	1	consistency	**0.82580511**
						kappa	**0.65205497**
cut1	**0.91641306**		cut2	**0.91851469**		cut3	**0.99086996**
0.10006185	0.04064691		0.76926147	0.04394656		0.98655872	0.00155148
0.04294003	0.81635122		0.03753875	0.14925322		0.00757855	0.00431124
	1			1			1

BEST COPY AVAILABLE

Step 4							
	Predicted Classification (X1)						
True Status	Fail	Needs	Prof	Adv	Marginal		
Fail	**0.16148386**	**0.03588105**	**4.0828E-06**	**9.7664E-12**	0.197369		
Needs	**0.05183143**	**0.38525032**	**0.05514207**	**0.00021392**	0.49243774		
Prof	**4.6587E-06**	**0.04205347**	**0.15715803**	**0.02108233**	0.22029849		
Adv	**2.0169E-12**	**8.9769E-05**	**0.02167632**	**0.06812869**	0.08989478		
Marginal	0.21331995	0.46327461	0.23398051	0.08942493	1		
Step 5							
	Actual Classification (X0)						
True Status	Fail	Needs	Prof	Adv	Marginal		
Fail	0.16675711	0.03394705	4.0337E-06	1.204E-11	0.2007082		
Needs	0.05352398	0.36448525	0.05447875	0.00026372	0.47275171		
Prof	4.8108E-06	0.03978678	0.15526754	0.02599056	0.22104968		
Adv	2.0828E-12	8.493E-05	0.02141556	0.08398992	0.10549041		
Marginal	0.2202859	0.43830402	0.23116589	0.11024419	1	accuracy	**0.77049981**
cut1	**0.91252012**		cut2	**0.90537697**		cut3	**0.95224523**
0.16675711	0.03395109		0.6187134	0.0547465		0.86825531	0.02625428
0.05352879	0.74576301		0.03987652	0.28666357		0.02150049	0.08398992
	1			1			1
Step 6							
	X1						
X2	Fail	Needs	Prof	Adv	Marginal		
Fail	0.15181376	0.06113341	0.00037251	2.7507E-07	0.21331995		
Needs	0.06113341	0.33414973	0.06654147	0.00145	0.46327461		
Prof	0.00037251	0.06654147	0.13875516	0.02831137	0.23398051		
Adv	2.7507E-07	0.00145	0.02831137	0.05966329	0.08942493		
Marginal	0.21331995	0.46327461	0.23398051	0.08942493	1		
Step 7							
	X0						
X2	Fail	Needs	Prof	Adv	Marginal		
Fail	0.15677123	0.0578383	0.00036803	3.3911E-07	0.2149779		
Needs	0.06312972	0.316139	0.06574102	0.00178758	0.44679731		
Prof	0.00038468	0.06295487	0.13708603	0.03490261	0.23532819		
Adv	2.8406E-07	0.00137185	0.0279708	0.07355366	0.10289659		
Marginal	0.2202859	0.43830402	0.23116589	0.11024419	1	consistency	**0.68354992**
						kappa	**0.54208463**
cut1	**0.87827865**		cut2	**0.86739135**		cut3	**0.93396654**
0.15677123	0.05820668		0.59387824	0.06789697		0.86041288	0.03669053
0.06351468	0.72150742		0.06471168	0.27351311		0.02934293	0.07355366
	1			1			1

Grade 4 Science and Technology.xls
Accuracy and Consistency of Classification

Step 4							
	Predicted Classification (X1)						
True Status	Fail	Needs	Prof	Adv	Marginal		
Fail	**0.07001338**	**0.01580572**	**1.1187E-06**	**4.7609E-14**	0.08582022		
Needs	**0.03189542**	**0.35986052**	**0.06498578**	**4.0151E-05**	0.45678187		
Prof	**4.9873E-06**	**0.06391559**	**0.30492396**	**0.03402635**	0.40287089		
Adv	**2.1538E-15**	**3.3368E-06**	**0.01250257**	**0.04202111**	0.05452702		
Marginal	0.10191379	0.43958517	0.38241343	0.07608761	1		
Step 5							
	Actual Classification (X0)						
True Status	Fail	Needs	Prof	Adv	Marginal		
Fail	0.074744	0.01441035	1.2519E-06	3.9081E-14	0.0891556		
Needs	0.03405051	0.32809117	0.07272646	3.2959E-05	0.43490109		
Prof	5.3243E-06	0.05827297	0.34124448	0.02793149	0.42745426		
Adv	2.2993E-15	3.0422E-06	0.01399179	0.03449421	0.04848905		
Marginal	0.10879983	0.40077753	0.42796399	0.06245866	1	accuracy	**0.77857386**
cut1	**0.95153257**		cut2	**0.868958**		cut3	**0.95804072**
0.074744	0.0144116		0.45129602	0.07276067		0.9235465	0.02796445
0.03405583	0.87678857		0.05828133	0.41766198		0.01399484	0.03449421
	1			1			1
Step 6							
	X1						
X2	Fail	Needs	Prof	Adv	Marginal		
Fail	0.06770548	0.03395013	0.00025814	3.2284E-08	0.10191379		
Needs	0.03395013	0.316304	0.0886404	0.00069063	0.43958517		
Prof	0.00025814	0.0886404	0.26060902	0.03290588	0.38241343		
Adv	3.2284E-08	0.00069063	0.03290588	0.04249107	0.07608761		
Marginal	0.10191379	0.43958517	0.38241343	0.07608761	1		
Step 7							
	X0						
X2	Fail	Needs	Prof	Adv	Marginal		
Fail	0.07228016	0.03095293	0.00028888	2.6502E-08	0.10352201		
Needs	0.03624405	0.28837992	0.09919866	0.00056693	0.42438955		
Prof	0.00027558	0.08081501	0.29165104	0.02701172	0.39975334		
Adv	3.4466E-08	0.00062966	0.03682541	0.03487999	0.0723351		
Marginal	0.10879983	0.40077753	0.42796399	0.06245866	1	consistency	**0.68719111**
						kappa	**0.51355663**
cut1	**0.93223849**		cut2	**0.81822522**		cut3	**0.93496623**
0.07228016	0.03124185		0.42785707	0.10005449		0.90008624	0.02757867
0.03651967	0.85995833		0.08172029	0.39036815		0.0374551	0.03487999
	1			1			1

Step 4							
	Predicted Classification (X1)						
True Status	Fail	Needs	Prof	Adv	Marginal		
Fail	**0.06009907**	**0.0254026**	**9.7898E-07**	**8.0799E-21**	0.08550265		
Needs	**0.04877865**	**0.41312695**	**0.05124401**	**5.1627E-09**	0.51314962		
Prof	**1.349E-06**	**0.04017139**	**0.27447534**	**0.01223793**	0.32688601		
Adv	**3.4646E-25**	**2.6983E-11**	**0.00676825**	**0.06769348**	0.07446173		
Marginal	0.10887907	0.47870094	0.33248857	0.07993141	1		
Step 5							
	Actual Classification (X0)						
True Status	Fail	Needs	Prof	Adv	Marginal		
Fail	0.07623671	0.01646419	1.5442E-06	2.7467E-21	0.09270244		
Needs	0.06187657	0.26775996	0.08083007	1.755E-09	0.4104666		
Prof	1.7112E-06	0.02603628	0.43294543	0.00416012	0.46314354		
Adv	4.395E-25	1.7488E-11	0.01067594	0.02301148	0.03368742		
Marginal	0.13811499	0.31026043	0.52445298	0.0271716	1	accuracy	**0.79995358**
cut1	**0.92165599**		cut2	**0.89313039**		cut3	**0.98516394**
0.07623671	0.01646573		0.42233742	0.08083161		0.96215246	0.00416012
0.06187828	0.84541928		0.026038	0.47079297		0.01067594	0.02301148
	1			1			1
Step 6							
	X1						
X2	Fail	Needs	Prof	Adv	Marginal		
Fail	0.05718146	0.05151397	0.00018364	2.3482E-12	0.10887907		
Needs	0.05151397	0.36308364	0.06410127	2.0641E-06	0.47870094		
Prof	0.00018364	0.06410127	0.25461634	0.01358732	0.33248857		
Adv	2.3482E-12	2.0641E-06	0.01358732	0.06634203	0.07993141		
Marginal	0.10887907	0.47870094	0.33248857	0.07993141	1		
Step 7							
	X0						
X2	Fail	Needs	Prof	Adv	Marginal		
Fail	0.07253567	0.03338775	0.00028967	7.9825E-13	0.10621309		
Needs	0.06534636	0.23532539	0.10111055	7.0167E-07	0.401783		
Prof	0.00023296	0.04154595	0.40162072	0.00461882	0.44801845		
Adv	2.9788E-12	1.3378E-06	0.02143204	0.02255208	0.04398546		
Marginal	0.13811499	0.31026043	0.52445298	0.0271716	1	consistency	**0.73203386**
						kappa	**0.57092**
cut1	**0.90074327**		cut2	**0.85681883**		cut3	**0.97394709**
0.07253567	0.03367742		0.40659517	0.10140092		0.95139502	0.00461953
0.06557932	0.82820759		0.04178025	0.45022366		0.02143338	0.02255208
	1			1			1

Step 4							
	Predicted Classification (X1)						
True Status	Fail	Needs	Prof	Adv	Marginal		
Fail	**0.37644546**	**0.04332445**	**0.00025316**	**3.4709E-09**	0.42002307		
Needs	**0.04597056**	**0.18850079**	**0.04389361**	**9.7548E-05**	0.2784625		
Prof	**0.00019034**	**0.03526247**	**0.1637655**	**0.02700418**	0.22622248		
Adv	**3.1806E-11**	**1.0403E-05**	**0.01361873**	**0.06166282**	0.07529195		
Marginal	0.42260635	0.2670981	0.221531	0.08876455	1		
Step 5							
	Actual Classification (X0)						
True Status	Fail	Needs	Prof	Adv	Marginal		
Fail	0.37381254	0.04251379	0.00026755	3.2893E-09	0.41659389		
Needs	0.04564903	0.18497371	0.04638966	9.2445E-05	0.27710485		
Prof	0.000189	0.03460266	0.17307817	0.02559135	0.23346119		
Adv	3.1584E-11	1.0208E-05	0.01439317	0.0584367	0.07284008		
Marginal	0.41965057	0.26210038	0.23412855	0.0841205	1	accuracy	**0.79030112**
cut1	**0.91138061**		cut2	**0.91844846**		cut3	**0.95991283**
0.37381254	0.04278135		0.64694907	0.04674967		0.90147613	0.0256838
0.04583804	0.53756808		0.03480188	0.27149939		0.01440338	0.0584367
	1			1			1
Step 6							
	X1						
X2	Fail	Needs	Prof	Adv	Marginal		
Fail	0.35950938	0.06047489	0.00261869	3.3894E-06	0.42260635		
Needs	0.06047489	0.1531248	0.05263184	0.00086657	0.2670981		
Prof	0.00261869	0.05263184	0.13831527	0.02796519	0.221531		
Adv	3.3894E-06	0.00086657	0.02796519	0.05992939	0.08876455		
Marginal	0.42260635	0.2670981	0.221531	0.08876455	1		
Step 7							
	X0						
X2	Fail	Needs	Prof	Adv	Marginal		
Fail	0.35699491	0.05934334	0.00276761	3.2121E-06	0.41910906		
Needs	0.06005192	0.15025965	0.05562479	0.00082124	0.2667576		
Prof	0.00260038	0.05164703	0.1461807	0.02650209	0.22693019		
Adv	3.3657E-06	0.00085036	0.02955546	0.05679396	0.08720315		
Marginal	0.41965057	0.26210038	0.23412855	0.0841205	1	consistency	**0.71022921**
						kappa	**0.58230453**
cut1	**0.87523018**		cut2	**0.88568201**		cut3	**0.94226428**
0.35699491	0.06211416		0.62664981	0.05921685		0.88547032	0.02732654
0.06265567	0.51823527		0.05510114	0.2590322		0.03040918	0.05679396
	1			1			1

Step 4							
	Predicted Classification (X1)						
True Status	Fail	Needs	Prof	Adv	Marginal		
Fail	**0.38965645**	**0.05864287**	**0.00061402**	**4.3263E-10**	0.44891334		
Needs	**0.05978038**	**0.20429412**	**0.04935458**	**1.3593E-05**	0.31344267		
Prof	**0.00035721**	**0.03314245**	**0.14970293**	**0.01356905**	0.19677164		
Adv	**5.1079E-12**	**8.093E-07**	**0.00624414**	**0.0346274**	0.04087235		
Marginal	0.44979405	0.29608025	0.20591566	0.04821004	1		
Step 5							
	Actual Classification (X0)						
True Status	Fail	Needs	Prof	Adv	Marginal		
Fail	0.34920697	0.06220993	0.0007882	1.6585E-10	0.4122051		
Needs	0.0535747	0.2167207	0.06335467	5.211E-06	0.33365527		
Prof	0.00032013	0.0351584	0.19216818	0.00520186	0.23284857		
Adv	4.5777E-12	8.5853E-07	0.00801537	0.01327484	0.02129106		
Marginal	0.4031018	0.31408989	0.26432641	0.01848191	1	accuracy	**0.77137068**
cut1	**0.88310705**		cut2	**0.90037253**		cut3	**0.9867767**
0.34920697	0.06299813		0.6817123	0.06414807		0.97350187	0.00520707
0.05389483	0.53390008		0.03547939	0.21866024		0.00801623	0.01327484
	1			1			1
Step 6							
	X1						
X2	Fail	Needs	Prof	Adv	Marginal		
Fail	0.36658882	0.07879105	0.00441317	9.9971E-07	0.44979405		
Needs	0.07879105	0.16270083	0.05437382	0.00021455	0.29608025		
Prof	0.00441317	0.05437382	0.13304172	0.01408695	0.20591566		
Adv	9.9971E-07	0.00021455	0.01408695	0.03390754	0.04821004		
Marginal	0.44979405	0.29608025	0.20591566	0.04821004	1		
Step 7							
	X0						
X2	Fail	Needs	Prof	Adv	Marginal		
Fail	0.32853395	0.08358366	0.00566503	3.8325E-07	0.41778302		
Needs	0.0706119	0.17259742	0.06979768	8.2251E-05	0.31308925		
Prof	0.00395505	0.05768121	0.17078079	0.0054004	0.23781745		
Adv	8.9594E-07	0.0002276	0.01808291	0.01299887	0.03131027		
Marginal	0.4031018	0.31408989	0.26432641	0.01848191	1	consistency	**0.68491102**
						kappa	**0.52958639**
cut1	**0.83618308**		cut2	**0.8625899**		cut3	**0.97620555**
0.32853395	0.08924908		0.65532693	0.07554535		0.96320669	0.00548304
0.07456785	0.50764913		0.06186476	0.20726297		0.01831141	0.01299887
	1			1			1

Step 4							
	Predicted Classification (X1)						
True Status	Fail	Needs	Prof	Adv	Marginal		
Fail	**0.34389855**	**0.04581933**	**1.5664E-05**	**6.284E-14**	0.38973355		
Needs	**0.04480894**	**0.24298113**	**0.03366964**	**2.5319E-06**	0.32146225		
Prof	**9.9662E-06**	**0.02624869**	**0.1487407**	**0.01481691**	0.18981627		
Adv	**4.5827E-16**	**2.4574E-07**	**0.00908125**	**0.08990644**	0.09898793		
Marginal	0.38871746	0.3150494	0.19150725	0.10472589	1		
Step 5							
	Actual Classification (X0)						
True Status	Fail	Needs	Prof	Adv	Marginal		
Fail	0.24763436	0.04927083	2.724E-05	2.8967E-14	0.29693243		
Needs	0.03226601	0.26128452	0.0585524	1.1671E-06	0.35210409		
Prof	7.1765E-06	0.02822596	0.25866399	0.00683004	0.29372718		
Adv	3.2999E-16	2.6425E-07	0.01579253	0.04144351	0.0572363		
Marginal	0.27990754	0.33878158	0.33303616	0.04827472	1	accuracy	**0.80902638**
cut1	**0.91842874**		cut2	**0.91318579**		cut3	**0.977376**
0.24763436	0.04929807		0.59045572	0.0585808		0.93593249	0.00683121
0.03227318	0.67079436		0.0282334	0.32273008		0.01579279	0.04144351
	1			1			1
Step 6							
			X1				
X2	Fail	Needs	Prof	Adv	Marginal		
Fail	0.324887	0.06333528	0.00049518	8.6971E-09	0.38871746		
Needs	0.06333528	0.20990919	0.04171874	8.6198E-05	0.3150494		
Prof	0.00049518	0.04171874	0.13244923	0.0168441	0.19150725		
Adv	8.6971E-09	8.6198E-05	0.0168441	0.08779558	0.10472589		
Marginal	0.38871746	0.3150494	0.19150725	0.10472589	1		
Step 7							
			X0				
X2	Fail	Needs	Prof	Adv	Marginal		
Fail	0.23394453	0.06810622	0.00086113	4.0091E-09	0.30291188		
Needs	0.04560645	0.22572132	0.07254999	3.9734E-05	0.34391748		
Prof	0.00035657	0.04486135	0.2303327	0.0077645	0.28331512		
Adv	6.2626E-09	9.2691E-05	0.02929234	0.04047048	0.06985552		
Marginal	0.27990754	0.33878158	0.33303616	0.04827472	1	consistency	**0.73046903**
						kappa	**0.61549041**
cut1	**0.88506963**		cut2	**0.88123854**		cut3	**0.96281072**
0.23394453	0.06896735		0.57337851	0.07345085		0.92234024	0.00780424
0.04596302	0.6511251		0.04531061	0.30786003		0.02938504	0.04047048
	1			1			1

BEST COPY AVAILABLE

Step 4							
	Predicted Classification (X1)						
True Status	Fail	Needs	Prof	Adv	Marginal		
Fail	**0.46948953**	**0.04035059**	**0.00013974**	**6.1866E-10**	0.50997986		
Needs	**0.04274758**	**0.17895177**	**0.03680915**	**6.8711E-05**	0.25857721		
Prof	**0.00012778**	**0.02964742**	**0.12676114**	**0.01970197**	0.17623831		
Adv	**9.3881E-11**	**1.5726E-05**	**0.01082169**	**0.0443672**	0.05520461		
Marginal	0.51236489	0.24896551	0.17453172	0.06413788	1		
Step 5							
	Actual Classification (X0)						
True Status	Fail	Needs	Prof	Adv	Marginal		
Fail	0.47197157	0.03969479	0.0001379	6.5376E-10	0.51180425		
Needs	0.04297357	0.17604333	0.03632364	7.261E-05	0.25541315		
Prof	0.00012845	0.02916557	0.12508916	0.02082003	0.17520321		
Adv	9.4377E-11	1.547E-05	0.01067895	0.04688497	0.05757939		
Marginal	0.5150736	0.24491915	0.17222964	0.06777761	1	accuracy	**0.81998903**
cut1	**0.91706528**		cut2	**0.93415636**		cut3	**0.96841294**
0.47197157	0.03983269		0.73068326	0.03653415		0.92152797	0.02089264
0.04310203	0.44509372		0.02930949	0.20347311		0.01069442	0.04688497
	1			1			1
Step 6							
			X1				
X2	Fail	Needs	Prof	Adv	Marginal		
Fail	0.45382116	0.05670352	0.00183829	1.9201E-06	0.51236489		
Needs	0.05670352	0.14711625	0.04442427	0.00072147	0.24896551		
Prof	0.00183829	0.04442427	0.10727156	0.02099759	0.17453172		
Adv	1.9201E-06	0.00072147	0.02099759	0.0424169	0.06413788		
Marginal	0.51236489	0.24896551	0.17453172	0.06413788	1		
Step 7							
			X0				
X2	Fail	Needs	Prof	Adv	Marginal		
Fail	0.45622037	0.05578193	0.00181405	2.0291E-06	0.51381837		
Needs	0.05700329	0.14472522	0.04383831	0.00076241	0.24632923		
Prof	0.00184801	0.04370226	0.10585665	0.02218918	0.17359609		
Adv	1.9303E-06	0.00070974	0.02072063	0.044824	0.0662563		
Marginal	0.5150736	0.24491915	0.17222964	0.06777761	1	consistency	**0.75162623**
						kappa	**0.61229514**
cut1	**0.88354876**		cut2	**0.90732126**		cut3	**0.95561408**
0.45622037	0.05759801		0.7137308	0.0464168		0.91079008	0.02295362
0.05885323	0.4273284		0.04626194	0.19359045		0.02143231	0.044824
	1			1			1

Grade 10 Science and Technology.xls
Accuracy and Consistency of Classification

Step 4							
	Predicted Classification (X1)						
True Status	Fail	Needs	Prof	Adv	Marginal		
Fail	**0.36076175**	**0.04375652**	**3.8481E-06**	**1.9702E-14**	0.40452212		
Needs	**0.0499724**	**0.30775337**	**0.03726004**	**7.8859E-06**	0.3949937		
Prof	**3.2579E-06**	**0.02699586**	**0.12696903**	**0.01200619**	0.16597434		
Adv	**9.8601E-16**	**8.925E-07**	**0.00607193**	**0.02832646**	0.03439928		
Marginal	0.41073741	0.37850664	0.17030485	0.04034054	0.99988943		
Step 5							
	Actual Classification (X0)						
True Status	Fail	Needs	Prof	Adv	Marginal		
Fail	0.30942463	0.04893441	4.7571E-06	6.7781E-15	0.35836379		
Needs	0.04286123	0.34417111	0.04606195	2.713E-06	0.433097		
Prof	2.7943E-06	0.03019039	0.15696285	0.00413056	0.19128659		
Adv	8.457E-16	9.9812E-07	0.0075063	0.00974531	0.01725261		
Marginal	0.35228865	0.42329691	0.21053586	0.01387859	1	accuracy	**0.8203039**
cut1	**0.90819681**		cut2	**0.9237364**		cut3	**0.98835943**
0.30942463	0.04893917		0.74539137	0.04606942		0.97861411	0.00413327
0.04286402	0.59877219		0.03019418	0.17834502		0.0075073	0.00974531
	1			1			1
Step 6							
	X1						
X2	Fail	Needs	Prof	Adv	Marginal		
Fail	0.34499023	0.06546389	0.00028328	9.3541E-09	0.41073741		
Needs	0.06546389	0.26794824	0.04492801	0.00016651	0.37850664		
Prof	0.00028328	0.04492801	0.11226982	0.01282374	0.17030485		
Adv	9.3541E-09	0.00016651	0.01282374	0.02735028	0.04034054		
Marginal	0.41073741	0.37850664	0.17030485	0.04034054	0.99988943		
Step 7							
	X0						
X2	Fail	Needs	Prof	Adv	Marginal		
Fail	0.29589743	0.0732105	0.0003502	3.2181E-09	0.36945813		
Needs	0.05614824	0.29965567	0.05554133	5.7284E-05	0.41140252		
Prof	0.00024297	0.05024453	0.13879125	0.00441182	0.19369057		
Adv	8.023E-09	0.00018621	0.01585308	0.00940948	0.02544877		
Marginal	0.35228865	0.42329691	0.21053586	0.01387859	1	consistency	**0.74375382**
						kappa	**0.60852547**
cut1	**0.87004807**		cut2	**0.89337747**		cut3	**0.97949159**
0.29589743	0.07356071		0.72491184	0.05594882		0.97008212	0.00446911
0.05639122	0.57415065		0.05067372	0.16846563		0.0160393	0.00940948
	1			1			1